

GenCore version 5.1.5  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 2, 2003, 11:21:20 ; Search time 423.993 Seconds

(without alignments)  
6795.341 Million cell updates/sec

Title: US-09-719-554-3\_COPY\_2502\_2600

Perfect score: 99  
Sequence: 1 tcacccctgagcgcaagc.....tgtrcaagaagcgaggaa 99

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

GenBank: 1: gb\_ba: 2: gb\_hg: 3: gb\_in: 4: gb\_om: 5: gb\_ov: 6: gb\_pat: 7: gb\_ph: 8: gb\_pl: 9: gb\_pr: 10: gb\_ro: 11: gb\_sy: 12: gb\_un: 13: gb\_vl: 14: em\_ba: 15: em\_fun: 16: em\_hum: 17: em\_in: 18: em\_mu: 19: em\_om: 20: em\_or: 21: em\_ov: 22: em\_ph: 23: em\_pat: 24: em\_pl: 25: em\_ro: 26: em\_sts: 27: em\_un: 28: em\_vl: 29: em\_hg\_inv: 30: em\_hg\_inv: 31: em\_hg\_inv: 32: em\_hg\_inv: 33: em\_hg\_inv: 34: em\_hg\_inv: 35: em\_hg\_inv: 36: em\_hg\_inv: 37: em\_hg\_inv: 38: em\_hg\_inv: 39: em\_hg\_inv: 40: em\_hg\_inv: 41: em\_hg\_inv:

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	99	100.0	10499	6	AX007980
2	99	100.0	56093	6	AX329572
3	99	100.0	56093	6	HSAC000064
4	99	100.0	149194	9	AC007566
5	95.8	96.8	40205	9	AF045450
6	95.8	96.8	183951	2	AC027500
7	94.2	95.2	104853	9	AC117444
8	94.2	95.2	142742	9	AF121782
9	94.2	95.2	164988	2	AC080070
10	94.2	95.2	169286	2	AC073066
11	94.2	95.2	169462	2	AC024033
12	94.2	95.2	177720	9	AC019193
13	94.2	95.2	186723	2	AC027752
14	94.2	95.2	340000	9	HS21C080
15	93.2	94.1	120450	9	AC093393
16	92.6	93.5	8339	9	AL162912
17	92.6	93.5	46340	6	AX008043
18	92.6	93.5	88502	2	AC021967
19	92.6	93.5	113189	9	AC007883
20	92.6	93.5	128468	9	AL139090
21	92.6	93.5	135873	2	AC012547
22	92.6	93.5	148724	2	AC010778
23	92.6	93.5	150955	9	AL356111
24	92.6	93.5	156818	9	AC104182
25	92.6	93.5	162887	9	AC097635
26	92.6	93.5	167366	2	AC021774
27	92.6	93.5	169747	9	AL353532
28	92.6	93.5	174228	2	AC016105
29	92.6	93.5	176879	2	AC016709
30	92.6	93.5	177163	9	HS134E15
31	92.6	93.5	187399	9	AC067942
32	92.6	93.5	192815	9	AC013444
33	92.6	93.5	199520	2	AC110420
34	92.6	93.5	219476	9	AC068945
35	91	91.9	47340	9	AC069387
36	91	91.9	112085	9	AC022132
37	91	91.9	112085	9	AC026712
38	91	91.9	123182	9	AC093675
39	91	91.9	123253	9	AC078963
40	91	91.9	129770	9	AC104684
41	91	91.9	136901	9	AC073626
42	91	91.9	138003	2	AC018346
43	91	91.9	139111	9	AC008768
44	91	91.9	163166	9	AL356632
45	91	91.9	164943	9	AP005272

#### ALIGNMENTS

RESULT 1  
AX007980  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
human.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
Perin, J.P., Rieger, F. and Alliel, P.M.  
Nucleic sequence and deduced protein sequence family with human  
endogenous retroviral motifs, and their uses

Pred. No. is the number of results predicted by chance to have a

JOURNAL Patent: WO 96/7395-A 3 29-DEC-1999;  
INST NAT SANTE RECH MED (FR); PERIN JEAN PIERRE (FR); RIEGER  
FRANCOIS (FR); ALLIEL PATRICK M (FR)  
FEATURES location/Qualifiers  
source 1..10499  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
BASE COUNT 3048 a 2676 c 2280 g 2495 t  
ORIGIN

Query Match 100.0%; Score 99; DB 6; Length 10499;  
Best Local Similarity 100.0%; Pred. No. 9.7e-24;  
Matches 99; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCACCCCTGAGCGCAAGGCAATGTTGGCATGCTGTTAAGAGCACCTAGATCCAGCA 60  
TCACCCCTGAGCGCAAGGCAATGTTGGCATGCTGTTAAGAGCACCTAGATCCAGCA 2561

Db 2502 TCACCCCTGAGCGCAAGGCAATGTTGGCATGCTGTTAAGAGCACCTAGATCCAGCA 2561

QY 61 GCCCAGACCCCTTCTTGTGTCAGAAAGCGCGGAAA 99  
GCCCAGACCCCTTCTTGTGTCAGAAAGCGCGGAAA 2600

Db 2562 GCCCAGACCCCTTCTTGTGTCAGAAAGCGCGGAAA 2600

RESULT 2  
AX329572 56093 bp DNA linear PAT 09-JAN-2002  
LOCUS  
DEFINITION Sequence 81 from Patent WO0194629.  
ACCESSION AX329572  
VERSION AX329572.1 GI:18102550  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1  
AUTHORS Young, P. E., Augustus, M., Carter, K. C., Ebner, R., Endress, G.,  
Horrigan, S., Soppet, D. R. and Weaver, Z.  
TITLE Cancer gene determination and therapeutic screening using signature  
gene sets  
JOURNAL Patent: WO 0194629-A 81 13-DEC-2001;  
Avalon Pharmaceuticals (US)  
FEATURES location/Qualifiers  
source 1..56093  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
BASE COUNT 16164 a 12346 c 10702 g 16881 t  
ORIGIN

Query Match 100.0%; Score 99; DB 6; Length 56093;  
Best Local Similarity 100.0%; Pred. No. 1.1e-23;  
Matches 99; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TCACCCCTGAGCGCAAGGCAATGTTGGCATGCTGTTAAGAGCACCTAGATCCAGCA 60  
TCACCCCTGAGCGCAAGGCAATGTTGGCATGCTGTTAAGAGCACCTAGATCCAGCA 30561

Db 30502 TCACCCCTGAGCGCAAGGCAATGTTGGCATGCTGTTAAGAGCACCTAGATCCAGCA 30561

QY 61 GCCCAGACCCCTTCTTGTGTCAGAAAGCGCGGAAA 99  
GCCCAGACCCCTTCTTGTGTCAGAAAGCGCGGAAA 30600

Db 30562 GCCCAGACCCCTTCTTGTGTCAGAAAGCGCGGAAA 30600

RESULT 3  
HSAC000064 56093 bp DNA linear PRI 13-NOV-1996  
LOCUS  
DEFINITION Human BAC clone RG083M05, from 7q21-7q22, complete sequence.  
ACCESSION AC000064  
VERSION AC000064.1 GI:1669369  
KEYWORDS HTG.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 56093)

AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

Pauley, A.  
The sequence of H. sapiens BAC clone RG083M05  
Unpublished (1996)  
2 (bases 1 to 56093)

Waterston, R.  
Direct Submission  
Submitted (13-NOV-1996)  
Genome Sequencing Center  
Department of Genetics, Washington University  
St. Louis, MO 63108, USA  
e-mail: sapiens@watson.wustl.edu

NOTICE: This sequence may not represent the entire insert of this  
clone. It may be shorter because we only sequence overlapping  
sections once, or longer because we provide a small overlap between  
neighboring submissions.

This sequence was finished as follows unless otherwise noted:  
all regions were double stranded or sequenced with an alternate  
chemistry; an attempt was made to resolve all sequencing problems,  
such as compressions and repeats; all regions were covered by  
sequence from more than one subclone; and the assembly was  
confirmed by restriction digest.

SOURCE INFORMATION:  
This clone is from the first release of the human BAC library. The  
library contains cloned DNA from a human male fibroblast cell line  
978SK. For references see: Shizuya et al., Proc. Natl. Acad. Sci.  
89:8794-8797 (1992); Kim et al., Genomics 34:213-218 (1996).  
VECTOR: pBEO  
Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:  
The orientation of this clone is unknown. Actual start of this  
clone is at base position 1 of H\_RG083M05; actual end is at 56093  
of H\_RG083M05

FEATURES  
source

This clone contains STS SMS1725.  
location/Qualifiers  
1..56093  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="7"  
/map="7q21-7q22"  
/clone\_lib="CITB-978SK-B"  
/clone\_1lb="CITB-978SK-B"  
complement(838..1131)  
/rpl\_family="ALU"  
/rpl\_16971  
/gene="WUGSC:H\_RG083M05.1"  
join<1360..1503,4181..4370,4587..4774,6422..6556,  
9483..9547,11631..11773,11864..12021,13131..13296,  
14885..14988,16349..16546,16837..16971)  
/gene="WUGSC:H\_RG083M05.1"  
/note="Atpase; strong similarity to peroxisome  
biogenesis protein PAB1 (PDB:1112019); coded for by  
human cDNA C04279 (NID:g1467530)"  
/codon\_start=1  
/protein\_id="AAB46346.1"  
/db\_xref="GI:1669371"  
/translation="KRLENIQKTLFEVASEAVMOPSVLLDLDLIALPAVPEHEH  
SPDARCEILCNVINKIKLDCDINKFTDDLDHAKETGFGVARDFTVADVAISRL  
SROSISTREKLVLTIDPQKALRGFLPALSRSVNLHKPPDLMDKIGTGHVROILMD  
TIOIPAKVCLKEKYEPELANPIROKCIILXGPGCGKTLIACVIARESMNFTSV  
KPELISKTIGSEDAVRDIFTRQAARFCILFFDEFESIAARRCHDNTGYDRVYNO  
LITOLDGVEGLGVVLAATSRPDLIDPALLRGRDKCYCPDPQVITVYLESTQV  
QMLHSELVSRLEILNVLSDPLADVDVQLHVAASYDSEFGADLALALYNADEALHG  
MLSKMSLELPDESERNRYLFGSEYESELSDSLSOCLSPSSMTODLGPV  
GKDOLESOPPVLRRTASOEGCOELTOEOROLRADISIIKGRYRSOGEBSMNQPGPI  
KTRLAISQSHLMTALGHTRPSTSEDDMKNFABL"

repeat\_region

gene

CDS

repeat\_region  
repeat\_region  
repeat\_region

complement(4948..5130)  
/rpl\_family="ALU"  
complement(6581..7133)

```

repeat_region      /rpt_family="LI1"
                   complement(7767, .8037)
repeat_region      /rpt_family="ALU"
                   complement(8186, .8472)
misc_feature        /rpt_family="ALU"
                   8473, .8625
                   /gene="WUGSC:H_RG083M05.1"
                   /note="match to human 3' EST H75782 (NID:g1049794), bases
                   287-444"
misc_feature        /gene="WUGSC:H_RG083M05.1"
                   8841, .9161
                   /note="match to human 5' EST H75921 (NID:g1050050), bases
                   21-348"
misc_feature        /gene="WUGSC:H_RG083M05.1"
                   9481, .9547
                   /note="match to human 5' EST N22627 (NID:g1130501), bases
                   276-343"
repeat_region      complement(12612, .12907)
misc_feature        /rpt_family="ALU"
                   13670, .13793
                   /gene="WUGSC:H_RG083M05.1"
                   /note="match to human 5' EST H41382 (NID:g17434), bases
                   143-266"
repeat_region      13794, .13877
misc_feature        /rpt_family="ALU"
                   13878, .13906
                   /gene="WUGSC:H_RG083M05.1"
                   /note="match to human 5' EST H41382 (NID:g17434), bases
                   30-58"
repeat_region      13907, .14104
                   /rpt_family="ALU"
                   complement(14110, .14137)
repeat_region      /rpt_family="LI"
                   complement(15618, .15907)
repeat_region      /rpt_family="ALU"
                   17227, .17522
misc_feature        /rpt_family="ALU"
                   18667, .19235
                   /note="match to human fetal brain 5' EST D61494
                   (NID:g970409), bases 1-255, and to human 3' EST R07476
                   (NID:g753399)"
repeat_region      19550, .19670
misc_feature        /rpt_family="ALU"
                   21507, .37303
                   /note="similarity to various SS-RNA virus polyproteins;
                   pseudogene; region of matches and close matches to
                   multiple human ESTs, see R68740 (NID:g842257)"
exon                37316, .37489
                   /note="Gtail prediction, score = 80"
repeat_region      /evidence="not_experimental"
                   complement(38938, .39224)
misc_feature        /rpt_family="ALU"
                   39225, .39707
                   /note="match to multiple human ESTs, see N30113
                   (NID:g114863)"
repeat_region      39800, .40085
                   /rpt_family="ALU"
                   complement(40247, .40538)
repeat_region      /rpt_family="ALU"
                   complement(40632, .40924)
repeat_region      /rpt_family="ALU"
                   complement(42283, .42891)
repeat_region      /rpt_family="ALU"
                   complement(45474, .45613)
misc_feature        /rpt_family="ALU"
                   complement(45614, .45737)
                   /note="match to human 3' EST H48898 (NID:g988738), bases
                   129-333"
misc_feature        complement(46107, .47026)
                   /note="match to multiple human ESTs, see N81064
                   (NID:g1243765), H48897 (NID:g988737), and M78831
                   (NID:g273146)"
repeat_region      complement(47027, .47318)

```

```

misc_feature      /rpt_family="ALU"/
                  complement(47365..47782)
                  /note="match to multiple human ESTs, see W37495
                  (NID:g1319089)"
misc_feature      47898..48115
                  /note="match to human 5' EST H62306 (NID:g1015138), bases
                  93-368"
repeat_region     complement(48116..48405)
                  /rpt_family="ALU"
misc_feature      complement(48406..48584)
                  /note="match to human 3' EST N29952 (NID:g1148472), bases
                  290-455, and 5' EST R12730 (NID:g765806)"
repeat_region     complement(48787..49405)
                  /rpt_family="ALU"
misc_feature      complement(49406..49534)
                  /note="match to human 3' EST R65794 (NID:g838432), bases
                  309-440"
repeat_region     complement(49638..49672)
                  /rpt_family="ALU"
misc_feature      complement(49674..49890)
                  /note="match to human 3' EST N29952 (NID:g1148472) and 5'
                  EST N29938 (NID:g1148458), sequences are from opposite
                  ends of the same clone"
gene              complement(49698..51806)
                  /gene="WUGSC:H_RG083M05.2"
                  complement(join(49698..49888,51575..51806))
                  /gene="WUGSC:H_RG083M05.2"
                  /note="coded for by human cDNAs W7389 (NID:g1319205),
                  R65891 (NID:g838529), R65794 (NID:g838432) and R65794
                  (NID:g838432)"
                  /codon_start=1
                  /protein_id="AAB46345.1"
                  /db_xref="GI:1669370"
                  /translation="MEFYEGCGIIFCEPGVYVOIGDVSVIDEODKPYVAIRGEF
                  IODQYCKSAATLWLIPILSPRPOPPASYSIGPEDLPKMKYLEFVCAPSEYRK
                  SRSPFFTVPLPRPKGYITWHVGTPTAITIKESVANHL"
exon              complement(51576..51756)
                  /gene="WUGSC:H_RG083M05.2"
                  /note="Grail prediction, score = 86"
                  /evidence=not experimental
                  complement(52052..52322)
                  /rpt_family="L1"
                  misc_feature      55557..55843
                  /note="match to human EST M79192 (NID:g273505) base 2-289"

Query Match      100.0%; Score 99; DB 9; Length 56093;
Best Local Similarity 100.0%; Pred. NO. 1.1e-23;
Matches          99; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY       1 TCACCCCTGAGGGCAAAAGCATGTTGGGCATGCGGTAAAGGACCCTAGATTCAGCA 60
         GCACCTTCTTGCTGTGTGTCACAAGAAAGCGGGGAAA 99
Db       30502 TCACCCCTGAGGGCAAAAGCATGTTGGGCATGCGGTAAAGGACCCTAGATTCAGCA 30561
         GCCCAGACCCCTTCTTGCTGTGTGTCACAAGAAAGCGGGGAAA 30600

RESULT 4
LOCUS      AC007566/c
DEFINITION Homo sapiens BAC clone CTB-1065 from 7q21-7q22, complete sequence.
VERSION    AC007566.2 GI:11181861
KEYWORDS   HTG.
SOURCE     Homo sapiens.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE  1 (bases 1 to 149194)
            Sulston,J.E. and Waterston,R.
            Toward a complete human genome sequence
            Genome Res. 8 (11), 1097-1108 (1998)
```

MEDLINE  
99063792  
PUBMED  
9847074  
2 (bases 1 to 149194)  
Du, Z.  
TITLE  
The sequence of Homo sapiens BAC clone CTB-10G5  
JOURNAL  
Unpublished (2001)  
REFERENCE  
3 (bases 1 to 149194)  
AUTHORS  
Waterston, R.H.  
TITLE  
Direct Submission  
JOURNAL  
Submitted (15-MAY-1999) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA  
REFERENCE  
4 (bases 1 to 149194)  
AUTHORS  
Waterston, R.  
TITLE  
Direct Submission  
JOURNAL  
Submitted (02-OCT-2000) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
REFERENCE  
5 (bases 1 to 149194)  
AUTHORS  
Waterston, R.H.  
TITLE  
Direct Submission  
JOURNAL  
Submitted (16-NOV-2000) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA  
REFERENCE  
6 (bases 1 to 149194)  
AUTHORS  
Waterston, R.H.  
TITLE  
Direct Submission  
JOURNAL  
Submitted (03-JAN-2002) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA  
REFERENCE  
7 (bases 1 to 149194)  
AUTHORS  
Waterston, R.H.  
TITLE  
Direct Submission  
JOURNAL  
Submitted (06-FEB-2002) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA  
REFERENCE  
8 (bases 1 to 149194)  
AUTHORS  
Waterston, R.  
TITLE  
Direct Submission  
JOURNAL  
Submitted (01-MAR-2002) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
On Nov 16, 2000 this sequence version replaced gi:4835815.  
----- Genome Center  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: <http://genome.wustl.edu/gsc>  
Contact: [saplens@wustl.wustl.edu](mailto:saplens@wustl.wustl.edu)  
----- Summary Statistics  
Center project name: H\_RG010605

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

#### MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/CTB/CHR7>, send <mailto:egreen@nhgri.nih.gov>, or see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:  
Clone CTB-10G5 is from the first release of the human BAC library CTB-978SK-B. The library contains cloned DNA from the male fibroblast cell line 978SK. See: Shizuya et al., Proc. Natl. Acad. Sci. USA 89:8794-7 (1992); U-J. Kim et al., Genomics 34:213-8 (1996). This clone is available from Research Genetics, Inc. (<http://www.resgen.com>).  
VECTOR: pBelobAC11  
Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:  
The clone sequenced to the left is RP5-911H5, 200 base pair overlap. Actual start of this clone is at base position 195 of CTB-10G5; actual end is at base position 150532 of CTB-10G5.

#### FEATURES

##### source

The clone CTB-10G5 contains the entire sequence of CTB-83M5.  
Location/Qualifiers  
1. .149194  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="7"  
/map="7q21-7q22"  
/clone="CTB-10G5"  
/clone\_11b="CTB-978SK-B"  
1. .1634  
/rpt\_family="L1"

repeat\_region  
2248. .2388  
/note="match to EST BG752883 (NID:g14063536)"

misc\_feature  
2248. .2387  
/note="match to EST AA149693 (NID:g1720635) zn99d06.r1"

misc\_feature  
2248. .2387  
/note="match to EST AW579261 (NID:g7254310)"

misc\_feature  
2248. .2387  
/note="match to EST BG766882 (NID:g14077335)"

misc\_feature  
2248. .2387  
/note="match to EST B1160365 (NID:g14620366)"

misc\_feature  
2248. .2374  
/note="similar to Homo sapiens EST BF758865 (NID:g12106765)"

misc\_feature  
2248. .2287  
/note="match to EST AV686676 (NID:g10288539)"

misc\_feature  
2253. .2387  
/note="match to EST AU123510 (NID:g10948226)"

misc\_feature  
2344. .2387  
/note="match to EST AV686676 (NID:g10288539)"

repeat\_region  
2696. .3066  
/rpt\_family="MaLR"

repeat\_region  
3108. .3392  
/rpt\_family="Alu"

misc\_feature  
3540. .3628  
/note="match to EST AA425526 (NID:g2106267) zw48b03.r1"

misc\_feature  
3540. .3628  
/note="similar to Mus musculus EST BB253526 (NID:g8946272)"

misc\_feature  
3542. .3632  
/note="match to EST AV686676 (NID:g10288539)"

misc\_feature  
3542. .3628  
/note="match to EST AA149693 (NID:g1720635) zn99d06.r1"

misc\_feature  
3542. .3628  
/note="match to EST AU123510 (NID:g10948226)"

misc\_feature  
3542. .3628  
/note="match to EST BG766882 (NID:g14077335)"

misc\_feature  
3542. .3628  
/note="match to EST B1160365 (NID:g14620366)"

misc\_feature  
3562. .3628  
/note="match to EST AW579261 (NID:g7254310)"

misc\_feature  
3598. .3628  
/note="match to EST BE272564 (NID:g9146913)"

misc\_feature  
3714. .3785  
/note="match to EST BE299708 (NID:g9183456)"

misc\_feature  
3717. .3785  
/note="match to EST BG260659 (NID:g12770475)"



/note="similar to Mus musculus EST BE994936  
(NID:g10678674)"  
misc\_feature 3719..3785  
/note="match to EST AA149693 (NID:g1720635) zn99d06.r1"  
misc\_feature 3719..3785  
/note="match to EST AA425526 (NID:g2106267) zw48b03.r1"  
misc\_feature 3719..3785  
/note="match to EST AU123510 (NID:g10948226)"  
misc\_feature 3719..3785  
/note="match to EST BE272564 (NID:g9146913)"  
misc\_feature 3719..3785  
/note="match to EST BE299708 (NID:g9183456)"  
misc\_feature 3719..3785  
/note="match to EST BG766882 (NID:g14077535)"  
misc\_feature 3719..3785  
/note="match to EST BI160365 (NID:g14620366)"  
misc\_feature 3719..3785  
/note="similar to Mus musculus EST BB253526  
(NID:g8946272)"  
misc\_feature 3860..3862  
/note="match to EST AV686676 (NID:g10288539)"  
repeat\_region 4730..4827  
/rpt\_family="L2"  
repeat\_region 4834..4925  
/rpt\_family="L1"  
misc\_feature 5715..5860  
/note="match to EST AA425526 (NID:g2106267) zw48b03.r1"  
misc\_feature 5715..5860  
/note="match to EST AU123510 (NID:g10948226)"  
misc\_feature 5715..5860  
/note="match to EST BE272564 (NID:g9146913)"  
misc\_feature 5715..5860  
/note="match to EST BE299708 (NID:g9183456)"  
misc\_feature 5715..5860  
/note="match to EST BG260659 (NID:g12770475)"  
misc\_feature 5715..5860  
/note="match to EST BG766882 (NID:g14077535)"  
misc\_feature 5715..5860  
/note="match to EST BI160365 (NID:g14620366)"  
misc\_feature 5715..5860  
/note="similar to Mus musculus EST BE994936"

Query Match 100.0%; Score 99; DB 9; Length 149194;  
Best Local Similarity 100.0%; Pred. No. 1,1e-23;  
Matches 99; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TCACCCCTGAGCGCAAGCAATGTTGGCATGCTGTAAGGACCACTAGATCCAGCA 60  
|||||  
Db 91502 TCACCCCTGAGCGCAAGCAATGTTGGCATGCTGTAAGGACCACTAGATCCAGCA 91443  
|||||

OY 61 GCCCAGACCCCTTCTTGTGTCAAGAAAGCGGGGAAA 99  
|||||  
Db 91442 GCCCAGACCCCTTCTTGTGTGTCAAGAAAGCGGGGAAA 91404  
|||||

RESULT 5  
AF045450/c 40205 bp DNA 1linear PRI 20-MAR-1998  
LOCUS Homo sapiens chromosome 21q22.3 cosmid Q11M15, complete sequence.  
ACCESSION AF045450.1 GI:2895783  
VERSION AF045450.1 GI:2895783  
KEYWORDS HTG.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 40205)  
AUTHORS Taudien,S., and Rosenthal,A.  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 40205)  
AUTHORS Taudien,S., Norstiek,G., Dagand,E., Hildmann,T., Drescher,B.,  
Weber,J., Rosenthal,A. and Yaspo,M.L.  
TITLE Direct Submission

JOURNAL Submitted (29-JAN-1998) Genome Analysis, Institut for Molecular  
Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany  
FEATURES  
Source  
1..40205  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="21"  
/map="21q22.3"  
/clone="cosmid Q11M15"  
complement(293..586)  
/evidence="not\_experimental"  
/rpt\_family="AluSc"  
complement(643..964)  
/evidence="not\_experimental"  
/rpt\_family="MERVL"  
complement(1053..1176)  
/note="Genscan, score = 5.19%, comment = Internal\_exon 124  
bp frame: 2 phase: 1"  
/evidence="not\_experimental"  
complement(1251..1859)  
/evidence="not\_experimental"  
/rpt\_family="HERVL"  
complement(2296..3220)  
/evidence="not\_experimental"  
/rpt\_family="HERV16"  
2818..2889  
/note="Xpound exon prediction, score = 82% (0%)"  
/evidence="not\_experimental"  
complement(3221..3658)  
/evidence="not\_experimental"  
/rpt\_family="MLTIC"  
3359..3395  
/note="Xpound exon prediction, score = 84% (0%)"  
/evidence="not\_experimental"  
complement(3874..4116)  
/evidence="not\_experimental"  
/rpt\_family="AluDb"  
complement(4330..4433)  
/evidence="not\_experimental"  
/rpt\_family="L2"  
4585..4638  
/note="GRAIL, score = 95.000%, comment = excellent shadow"  
/evidence="not\_experimental"  
complement(5113..7899)  
/evidence="not\_experimental"  
/rpt\_family="LIPAL4"  
complement(6119..6598)  
/note="GRAIL, score = 63.000%, comment = good"  
/evidence="not\_experimental"  
complement(7927..8045)  
/evidence="not\_experimental"  
/rpt\_family="FLAM\_A"  
8048..8178  
/evidence="not\_experimental"  
/rpt\_family="L2"  
8771..8876  
/note="WZEF, score = 67.4%"  
/evidence="not\_experimental"  
complement(8809..9187)  
/evidence="not\_experimental"  
/rpt\_family="MER1A"  
8837..8876  
/note="GRAIL, score = 55.000%, comment = good shadow"  
/evidence="not\_experimental"  
complement(9113..9213)  
/note="GRAIL, score = 55.000%, comment = good"  
/evidence="not\_experimental"  
complement(9190..9258)  
/evidence="not\_experimental"  
/rpt\_family="L1MC3"  
9259..9647  
/evidence="not\_experimental"  
/rpt\_family="L1MC3"

```
repeat_region complement(10138, .10427)
/evidence-not_experimental
/rpt_family="MLT2E"
repeat_region 10662, .10875
/evidence-not_experimental
/rpt_family="MER58A"
exon complement(11072, .11156)
/note="GRAIL, score = 86.000%, comment = excellent"
/evidence-not_experimental
complement(11072, .11129)
exon /note="Xpound exon prediction, score = 89% (0%)"
/evidence-not_experimental
11478, .11597
repeat_region /evidence-not_experimental
/rpt_family="L2"
11570, .11704
/exon /note="MZF, score = 66.1%"
/evidence-not_experimental
11570, .11704
exon /note="GRAIL, score = 99.000%, comment = excellent shadow"
complement(11687, .11735)
repeat_region /evidence-not_experimental
/rpt_family="MTR"
complement(11774, .12320)
repeat_region /evidence-not_experimental
/rpt_family="MLT1E"
complement(11828, .11949)
exon /note="GRAIL, score = 50.000%, comment = good"
/evidence-not_experimental
12623, .12843
repeat_region /evidence-not_experimental
/rpt_family="AluSx"
13102, .13457
repeat_region /evidence-not_experimental
/rpt_family="THE1B"
complement(13481, .13519)
exon /note="GRAIL, score = 65.000%, comment = good"
/evidence-not_experimental
15065, .15188
repeat_region /evidence-not_experimental
/rpt_family="HERVL"
15130, .15378
/misc_feature /note="GC score = 10.60 (249bp); Region: GC content"
15369, .16268
repeat_region /evidence-not_experimental
/rpt_family="MERVL"
complement(16623, .16714)
exon /note="GRAIL, score = 72.000%, comment = good"
/evidence-not_experimental
16661, .16730
exon /note="MZF, score = 93.5%"
/evidence-not_experimental
16880, .17318
repeat_region /evidence-not_experimental
/rpt_family="MLT1C"
17326, .17436
repeat_region /evidence-not_experimental
/rpt_family="LTRL16C"
complement(17540, .17675)
repeat_region /evidence-not_experimental
/rpt_family="MIR"
complement(17805, .17864)
repeat_region /evidence-not_experimental
/rpt_family="MLT1C"
18076, .18098
exon /note="GRAIL, score = 45.000%, comment = marginal shadow"
/evidence-not_experimental
complement(18219, .18349)
exon /note="Genscan, score = 2.99%, comment = Internal_exon 131
bp frame: 1 phase: 2"
/evidence-not_experimental
```

```
repeat_region complement(18717, .19211)
/evidence-not_experimental
/rpt_family="MLT1C"
exon complement(19252, .19399)
/note="GRAIL, score = 71.000%, comment = good"
/evidence-not_experimental
19269, .19350
exon /note="GRAIL, score = 81.000%, comment = excellent shadow"
/evidence-not_experimental
complement(19379, .19399)
exon /note="Xpound exon prediction, score = 60% (0%)"
/evidence-not_experimental
complement(19501, .19881)
exon /note="MZF, score = 50.6%"
/evidence-not_experimental
19945, .20020
repeat_region /note="homology = 100.00%, score = 38, counts = 2"
/evidence-not_experimental
/rpt_type=tandem
/rpt_unit=ttcttgcacacttcctgttgagagtcgattatcgc
repeat_region 20223, .20643
/evidence-not_experimental
/rpt_family="MLT1E"
20645, .20910
repeat_region /evidence-not_experimental
/rpt_family="MLT1F"
20961, .21007
repeat_region /evidence-not_experimental
/rpt_family="MLT1F"
21483, .21781
repeat_region /evidence-not_experimental
/rpt_family="AluJb"
complement(21504, .21756)
exon /note="MZF, score = 79.4%"
/evidence-not_experimental
complement(22325, .22806)
repeat_region /evidence-not_experimental
/rpt_family="MLT1C"

Query Match 96.8%; Score 95.8; DB 9; Length 40205;
Best Local Similarity 98.0%; Pred. No. 14e-22;
Matches 97; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 TCACCCCTGAGCGCAAGGCAATGTTGGCATCTGTAAAGCACCCTAGATCCAGCA 60
DB 32980 TCACCCCTGAGCGCAAGGCAATGTTGGCATCTGTAAAGCACCCTAGATCCAGCA 32921
QY 61 GCCCAGACCCCTTCTTGTGTGCTCAAGAAAGCGCGGAAA 99
DB 32920 GCCCGAGACCCCTTCTTGTGTGCTCAAGAAAGCGCGGAAA 32882

RESULT 6
AC027500 183951 bp DNA linear HTG 27-APR-2000
LOCUS Homo sapiens chromosome 3 clone RP11-705F24 map 3, WORKING DRAFT
SEQUENCE, 15 unordered pieces.
ACCESSION AC027500
VERSION AC027500.2 GI:7651959
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 183951)
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.
TITLE Homo sapiens chromosome 3, clone RP11-705F24
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 183951)
REFERENCE Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,
Boguslavsky,L., Bouknight,B., Brown,A., Burkett,G.,
Campiolo,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,
```

TITLE  
JOURNAL  
COMMENT

Collamore, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S.,  
Dodgson, S., Domingo, M., Doyle, M., Ferreira, P., Fitzhugh, W., Gage, D.,  
Gadigan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,  
Grand-Pierre, N., Grant, G., Hegos, B., Heaford, A., Horton, L.,  
Holland, J. C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,  
Klein, J., Lacombe, K., Lamazares, R., Landers, T., Lehoczy, J.,  
Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N.,  
McCarthy, M., McKean, P., McGurk, A., McKernan, K., McPheters, R.,  
Meldrum, J., Meneus, L., Minova, T., Miranda, C., Mlenga, V., Morrow, J.,  
Murphy, T., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P.,  
O'Neill, D., Oliver, T. M., Oliver, J., Peterson, K., Pierre, N.,  
Pisan, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,  
Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,  
Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,  
Teafaye, S., Theodore, J., Tirrell, A., Travers, M., Triggillo, J.,  
Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J.,  
Young, G., Zainoun, J., Zimmer, A. and Zody, M.

Direct Submission

Submitted (30-MAR-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA

On Apr 27, 2000 this sequence version replaced gi:7342244.

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

Project Information

Center project name: L8456

Center clone name: 705\_F\_24

Summary Statistics

Sequencing vector: M13; M7815; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 175817 bases at least Q40  
Consensus quality: 180619 bases at least Q30  
Consensus quality: 181937 bases at least Q20  
Insert size: 18500; agarose-fp  
Insert size: 182551; sum-of-ctnigs  
Quality coverage: 4.9 in Q20 bases; agarose-fp  
Quality coverage: 4.9 in Q20 bases; sum-of-ctnigs

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 15 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

```

1. 1062: contig of 1062 bp in length
* 1063 1162: gap of 100 bp
* 1163 2742: contig of 1580 bp in length
* 2743 2842: gap of 100 bp
* 2843 4245: contig of 1403 bp in length
* 4246 4345: gap of 100 bp
* 4346 6301: contig of 1956 bp in length
* 6302 6401: gap of 100 bp
* 6402 9026: contig of 2625 bp in length
* 9027 9126: gap of 100 bp
* 9127 13381: contig of 4255 bp in length
* 13382 13481: gap of 100 bp
* 13482 19288: contig of 5807 bp in length
* 19289 19388: gap of 100 bp
* 19389 25973: contig of 6585 bp in length
* 25974 26073: gap of 100 bp
* 26074 35483: contig of 9416 bp in length
* 35490 35589: gap of 100 bp
* 35590 46220: contig of 10631 bp in length
* 46221 46320: gap of 100 bp
* 46321 62888: contig of 16568 bp in length

```

```

* 62889 62988: gap of 100 bp
* 62989 81495: contig of 18507 bp in length
* 81496 81595: gap of 100 bp
* 81596 99524: contig of 17925 bp in length
* 99525 99624: gap of 100 bp
* 99625 129294: contig of 29670 bp in length
* 129295 129394: gap of 100 bp
* 129395 183951: contig of 54557 bp in length.

```

Location/Qualifiers

Source

```

1. 183951
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3"
/map="3"
/clone="RP11-705P24"
/clone_lib="RP11 Human Male BAC"
1. 1062

```

misc\_feature

```

/note="assembly-fragment"
1163. 2742

```

misc\_feature

```

/note="assembly-fragment"
2843. 4245

```

misc\_feature

```

/note="assembly-fragment"
4346. 6301

```

misc\_feature

```

/note="assembly-fragment"
6402. 9026

```

misc\_feature

```

/note="assembly-fragment"
9127. 13381

```

misc\_feature

```

/note="assembly-fragment"
13482. 19288

```

misc\_feature

```

/note="assembly-fragment"
clone_end:77
vector_side:right"
19389. 25973

```

misc\_feature

```

/note="assembly-fragment"
26074. 35489

```

misc\_feature

```

/note="assembly-fragment"
clone_end:56
vector_side:left"
35590. 46220

```

misc\_feature

```

/note="assembly-fragment"
46321. 62888

```

misc\_feature

```

/note="assembly-fragment"
62989. 81495

```

misc\_feature

```

/note="assembly-fragment"
81596. 99524

```

misc\_feature

```

/note="assembly-fragment"
99625. 129294

```

misc\_feature

```

/note="assembly-fragment"
129395. 183951

```

misc\_feature

```

/note="assembly-fragment"

```

```

BASE COUNT 58152 a 35325 c 34804 g 54268 t 1402 others
ORIGIN

```

```

Query Match 96.8%; Score 95.8; DB 2; Length 183951;
Best Local Similarity 98.0%; Pred. No.1.5e-22;
Matches 97; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

QY 1 TCACCCCTGAGCGCAAGGCAATGTTGGCATGCTGCTGAAGACCACTAGATCCAGCA 60
|||||
DB 30861 TCACCCCTGAGCGCAAGGCAATGTTGGCATGCTGCTGAAGACCACTAGATCCAGCA 30920
|||||
QY 61 GCCCAGACCCCTTCTTGTGTGTCAGAAAGCGGGGAAA 99
|||||
DB 30921 GCCCAGACCCCTTCTTGTGTGTCAGAAAGCGGGGAAA 30959
|||||

```

RESULT 7

AC117444/c

LOCUS

DEFINITION

AC117444

ACCESSION

VERSION

```

AC117444 104853 bp DNA linear PRI 31-JUL-2002
Homo sapiens 3 BAC RP11-215L17 (Roswell Park Cancer Institute Human
BAC Library) complete sequence.
AC117444
AC117444.6 GI:22024337

```

KEYWORDS  
SOURCE  
ORGANISMREFERENCE  
AUTHORSHTG.  
human.Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 104853)

Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,  
Alsbrooks,S.L., Amaralung,H.C., Are,J.R., Ayele,M., Banks,T.,  
Barbata,J., Benton,J., Bimaga,K., Blankenburg,K., Bonnin,D.,  
Buhay,J., Bowie,S., Brieve,M., Brown,E., Brown,M., Bryant,N.P.,  
Carton,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,  
Chen,G., Chen,R., Chen,Z., Chowdhury,I., Christopoulos,C.,  
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,  
DeLaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,  
Doutharte,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,  
Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,  
Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,  
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,  
Gorrell,J.H., Guevara,M., Gunaratne,P., Hale,S., Hamilton,K.,  
Hernandez,J., Hernandez,O., Hodgson,A., Hoques,M., Holloway,C.,  
Hollins,B., Homai,F., Howard,S., Huber,J., Hulik,S., Hume,J.,  
Joudan,S., Karlsson,E., Kelly,S., Khan,U., Kling,L., Korvah,J.,  
Kovar,C., Kratoch,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C.,  
Lewis,L., Li,J., Li,Z., Licharge,O., Lieu,C., Liu,J., Liu,W.,  
Loulsege,H., Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R.,  
Me,J., Meshwari,M., Mapua,P., Martin,R., Martindale,A.,  
Martinez,E., Massey,E., Mawhney,E., McLeod,M.P., Meador,M.,  
Mel,G., Metzger,M., Miner,G., Miner,Z., Mitchell,T., Mohabbat,K.,  
Moore,S., Morgan,M., Moorish,T., Morris,S., Moser,M., Neal,D.,  
Nelson,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N., Nguyen,N.,  
Nickerson,E., Nwokkwo,S., Ogumy, Okunnu,G., Oragunye,N.,  
Orlando,R., Pace,A., Payton,B., Peery,J., Perez,L., Peters,L.,  
Pickens,R., Primus,E., Pu,L., Quiles,M., Ren,Y., Rives,M.,  
Rojas,A., Roubokan,I., Rolfe,M., Ruiz,S., Saverio,G., Scherer,S.,  
Scott,G., Shen,H., Shooshari,N., Sisson,I., Sodergren,E.,  
Sonakke,T., Sparks,A., Stanley,H., Stone,H., Sutton,A., Svatek,A.,  
Tabors,P., Tamerisa,A., Tamerisa,K., Tang,H., Tansey,J., Taylor,C.,  
Taylor,T., Telford,B., Thomas,N., Thomas,S., Usmani,K., Vasquez,L.,  
Vera,V., Villalón,D., Vinson,R., Wang,Q., Wang,S., Ward-Moore,S.,  
Warren,R., Washington,C., Watlington,S., Williams,G.,  
Williamson,A., Wleczky,R., Wooden,S., Worley,K., Wu,C., Wu,Y.,  
Wu,Y.F., Zhou,J., Zorrilla,S., Naylor,S.L., Weinstein,G. and  
Gibbs,R.

TITLE  
JOURNALUnpublished  
2 (bases 1 to 104853)REFERENCE  
AUTHORS  
JOURNALTITLE  
JOURNALREFERENCE  
AUTHORS  
JOURNALREFERENCE  
AUTHORS  
JOURNALREFERENCE  
AUTHORS  
JOURNAL

## COMMENT

CLONE LENGTH: This sequence does not necessarily represent the  
entire insert of this clone. Overlapping regions of clones are only

sequenced and submitted once, so the sequence for the remainder of  
the insert may be found in the record for the adjacent clones.  
Overlapping clones are noted at the beginning and end of the  
features listing.

## ANNOTATION OF FEATURES:

STS are identified using ePCR (Genome Res. 7:541-550) searches  
of a local database that includes entries from dbSTS, GDB, and  
local mapping efforts.  
Repeats are identified using RepeatMasker (A. Smit and P. Green,  
unpublished.) for Human and Mouse sequences.  
Genes and Region of sequence similarity are identified by BLAST  
(Nuc. Acids Res. 25:3389-3402) similarity (expect < 1e-34) to the  
EST and cDNA sequences. Genes demonstrate at least two exons  
flanked by consensus splice sites that maintained sequence  
continuity across the splice junctions. Sequences that are not  
identical matches are annotated as similar.

SEQUENCING READ COVERAGE: Sequencing is completed to a minimum  
standard of double strand coverage with a minimum of 2 clones and 2  
reads with no ambiguities or 2 chemistries with a minimum of 2  
clones and 3 reads with no ambiguities. If the sequence quality for  
a region does not meet this standard, it will be indicated in the  
annotation as Low Coverage.

QUALITY OF INDIVIDUAL BASES: This sequence meets stringent quality  
standards - estimated error rate less than 1 per 10,000 bases.  
Reports of lowest quality individual bases and measures of base  
quality are listed below. Description of the metrics can be found  
at URL: [http://www.hgsc.bcm.tmc.edu:8088/quality/info/genbank\\_annotation.html](http://www.hgsc.bcm.tmc.edu:8088/quality/info/genbank_annotation.html).

## FEATURES

## source

1. 104853

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="3"

/clone="RP11-215L17"

1. 2000

/note="overlaps bases 154275..156277 of clone AC107015"

/function="clone overlap"

246..266

/rpt\_family="(A)n"

411..954

/rpt\_family="MLT2A"

1576..1598

/rpt\_family="AT\_rich"

complement(2562..3082)

/rpt\_family="MLT1G"

4553..4922

/rpt\_family="MLT1A1"

5010..5035

/rpt\_family="AT\_rich"

5079..5406

/rpt\_family="L2"

6323..6351

/rpt\_family="(A)n"

6948..7099

/rpt\_family="MIR"

7338..7447

/rpt\_family="MER94"

8267..8294

/rpt\_family="AT\_rich"

9151..9242

/rpt\_family="MLT1J"

9286..9509

/rpt\_family="MLT1J2"

10028..10091

/rpt\_family="(CATATA)n"

complement(13636..14644)

/rpt\_family="MER45B"

15307..15335

/rpt\_family="AT\_rich"

15593..15628

```
repeat_region      /rpt_family="AT_rich"
15747..16002
/rpt_family="AluY"
repeat_region      16122..16353
/rpt_family="MIR"
repeat_region      17470..17502
/rpt_family="AT_rich"
complement(18294..18632)
/rpt_family="L1M4"
repeat_region      19259..19775
/rpt_family="L1PA13"
20757..20815
/rpt_family="CT_rich"
complement(21111..27252)
/rpt_family="L1PA5"
22521..22798
/standard_name="D11S3206"
STS
repeat_region      27253..27294
/rpt_family="AT_rich"
complement(27524..28145)
/rpt_family="L1MC1"
28155..28203
/rpt_family="AT_rich"
complement(28585..28682)
/rpt_family="U6"
29288..29952
/rpt_family="MER76"
30941..30960
/rpt_family="(TTTG)n"
complement(30963..31250)
/rpt_family="AluY"
repeat_region      complement(31825..31911)
/rpt_family="L2"
complement(32569..32880)
/rpt_family="AluSg"
33002..33226
/rpt_family="LTR59"
33247..33731
/rpt_family="MER72"
complement(33796..34214)
/rpt_family="L2"
complement(34233..35306)
/rpt_family="MER11C"
35706..35894
/rpt_family="MIR"
36208..36244
/rpt_family="AT_rich"
38541..38824
/rpt_family="AluX"
repeat_region

Query Match      95.2% Score 94.2 DB 9 Length 104853;
Best Local Similarity 97.0% Pred. No.5.3e-22;
Matches 96; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TCACCCCTGAGCAAGCAATGTTGGCATGCTGTAAGACCACTAGATCCAGCA 60
DB 92005 TCACCCCTGAGCAAGCAATGTTGGCATGCTGTAAGACCACTAGATCCAGCA 91946
QY 61 GCCCAGACCCCTTTCTTGTGTGTCAGAAAGCGCGGAAA 99
DB 91945 GCCCAGACCCCTTTCTTGTGTGTCAGAAAGCGCGGAAA 91907

RESULT 8
LOCUS AF121782/c AF121782 142742 bp DNA linear PRI 02-FEB-1999
DEFINITION Homo sapiens chromosome 21q22.3 PAC 206A10, complete sequence.
ACCESSION AF121782
VERSION AF121782.1 GI:4210991
KEYWORDS HMG. sapiens.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

---

```
REFERENCE          Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 142742)
AUTHORS            Taudien,S., Dagand,E., Hildmann,T., Nordsiek,G., Drescher,B.,
Schattneyov,R., Weber,J., Schilling,M., Menzel,U., Yaspo,M.L. and
Rosenthal,A.
TITLE              Direct Submission
JOURNAL            Submitted (21-JAN-1999) Genome Analysis, Institute for Molecular
Biotechnology, Beutenbergstrasse 11, Jena 07745, Germany
FEATURES
source
1..142742
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.3"
/clone="PAC 206A10"
3..118
/evidence-not_experimental
/rpt_family="MIR"
complement(140..415)
/evidence-not_experimental
/rpt_family="MLT2B"
complement(668..881)
/evidence-not_experimental
/rpt_family="MER58A"
complement(1078..1162)
/note="GRAIL, score = 86.000%, comment = excellent shadow"
1114..1403
/evidence-not_experimental
/rpt_family="MLT11"
1481..1603
/evidence-not_experimental
/rpt_family="L2"
1576..1710
/note="WZEF, score = 66.1%"
/evidence-not_experimental
1576..1710
/note="GRAIL, score = 99.000%, comment = excellent"
complement(1693..1741)
/evidence-not_experimental
/rpt_family="MIR"
complement(1780..2326)
/evidence-not_experimental
/rpt_family="MLT1E"
complement(1834..1955)
/note="GRAIL, score = 50.000%, comment = good shadow"
2629..2831
/evidence-not_experimental
/rpt_family="AluX"
3012..3122
/evidence-not_experimental
/rpt_family="MLT1J"
3123..3478
/evidence-not_experimental
/rpt_family="THE1B"
3479..3618
/evidence-not_experimental
/rpt_family="MLT1J"
complement(3502..3540)
/note="GRAIL, score = 65.000%, comment = good shadow"
4674..4781
/note="GRAIL, score = 42.000%, comment = marginal"
4712..4749
/note="Xpound exon prediction, score = 70% (0%)"
5385..6104
/evidence-not_experimental
/rpt_family="HERVL"
6682..6751
exon
repeat_region
exon
```

```
/note="MZEf, score = 95.1%"
/evidence-not_experimental
6746. .6751
/note="GenScan, score = 1.83%, comment = Initial_exon 6 bp
frame: 1 phase: 0"
/evidence-not_experimental
6901. .7339
/evidence-not_experimental
/rpt_family="MLT1C"
7347. .7457
/evidence-not_experimental
/rpt_family="LTP16C"
complement(7561. .7696)
/evidence-not_experimental
/rpt_family="MIR"
complement(7819. .8220)
/evidence-not_experimental
/rpt_family="MLT1J"
8097. .8119
/note="GRAIL, score = 45.000%, comment = marginal"
/evidence-not_experimental
complement(8738. .9232)
/evidence-not_experimental
/rpt_family="MLT1C"
complement(9273. .9420)
/note="GRAIL, score = 71.000%, comment = good"
/evidence-not_experimental
9290. .9371
/note="GRAIL, score = 81.000%, comment = excellent shadow"
/evidence-not_experimental
9666. .10041
/note="homology = 100.00%, score = 38, counts = 2"
/evidence-not_experimental
/rpt_type=tandem
/rpt_unit=ctctctgcacactctctgttgagaggtcggttatcgc
10244. .10664
/evidence-not_experimental
/rpt_family="MLT1E"
10665. .11028
/evidence-not_experimental
/rpt_family="MLT1F"
11318. .11417
/evidence-not_experimental
/rpt_family="L1MC/D"
11503. .11818
/evidence-not_experimental
/rpt_family="AluSg"
complement(11525. .11777)
/note="MZEf, score = 79.4%"
/evidence-not_experimental
complement(12346. .12827)
/evidence-not_experimental
/rpt_family="MLT1C"
12902. .13180
/evidence-not_experimental
/rpt_family="AluSg"
complement(13057. .13161)
/note="MZEf, score = 76.1%"
/evidence-not_experimental
complement(13226. .13364)
/evidence-not_experimental
/rpt_family="L1MC2"
complement(13613. .13908)
/evidence-not_experimental
/rpt_family="AluSx"
complement(14317. .14614)
/evidence-not_experimental
/rpt_family="L1R37A"
complement(14962. .15036)
/note="MZEf, score = 93%"
/evidence-not_experimental
15235. .15299
/evidence-not_experimental
```

```
/rpt_family="L1PA6"
complement(15469. .15497)
/note="Xpound exon prediction, score = 62% (0%)"
/evidence-not_experimental
16045. .16768
/evidence-not_experimental
/rpt_family="MER21A"
complement(16555. .16655)
/note="GRAIL, score = 54.000%, comment = good"
/evidence-not_experimental
16823. .16850
/evidence-not_experimental
/rpt_family="MER34"
complement(16851. .17208)
/evidence-not_experimental
/rpt_family="THB1A"
17209. .17632
/evidence-not_experimental
/rpt_family="MER34"
complement(17633. .17969)
/evidence-not_experimental
/rpt_family="AluSg1"
complement(18004. .18334)
/evidence-not_experimental
/rpt_family="L1R17"
complement(18335. .18632)
/evidence-not_experimental
/rpt_family="HERV17"
18352. .18420
/note="GRAIL, score = 51.000%, comment = good shadow"
/evidence-not_experimental
complement(18397. .18584)
/note="GRAIL, score = 66.000%, comment = good"
/evidence-not_experimental
complement(18643. .24647)
/evidence-not_experimental
/rpt_family="HERV17"
complement(19228. .19361)
/note="MZEf, score = 92.8%"
/evidence-not_experimental
20900. .21135
/note="MZEf, score = 51.4%"
/evidence-not_experimental
21150. .21275
/note="Xpound exon prediction, score = 74% (0%)"
/evidence-not_experimental
```

Query Match 95.2%; Score 94.2; DB 9; Length 142742;  
Best Local Similarity 97.0%; Pred. No. 5.3e-22;  
Matches 96; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TCCACCTGAGCCCAAGGCAATGTTGGCATGCTGTAAAGGACCACTGAAATCCAGCA 60  
DB 22987 TCACCTCTGAGCCCAAGGCAATGTTGGCATGCTGTAAAGGACCACTGAAATCCAGCA 22928

QY 61 GCCCAGACCCCTTCTTGTGTCAAGAAAGCGGGGAAA 99  
DB 22927 GCCCGAGCCCTTCTTGTGTGTCAAGAAAGCGGGGAAA 22889

RESULT 9  
AC080070 164988 bp DNA 11near HTG 05-NOV-2000  
LOCUS Homo sapiens chromosome 4 clone RP11-162D2, WORKING DRAFT SEQUENCE,  
DEFINITION 24 unoriented pieces.  
ACCESSION AC080070  
VERSION AC080070.2 GI:11079592  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 164988)

AUTHORS Waterston,R.H.  
TITLE The sequence of Homo sapiens clone  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 164988)  
AUTHORS Waterston,R.H.  
TITLE Direct Submission  
JOURNAL Submitted (23-SEP-2000) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

COMMENT On Nov 3, 2000 this sequence version replaced gi:10280941.

----- Genome Center -----  
Center: Washington University Genome Sequencing Center  
Center code: WTGSC  
Web site: http://genome.wustl.edu/gsc/index.shtml  
----- Project Information -----  
Center project name: H\_NH0162D2  
----- Summary Statistics -----  
Sequencing vector: plasmid; 100%  
Chemistry: Dye-Primer ET; 100% of reads  
Chemistry: Dye-terminator Big Dye; 0% of reads  
Assembly program: Phrap; version 0.990319  
Consensus quality: 150371 bases at least Q40  
Consensus quality: 155395 bases at least Q30  
Consensus quality: 157776 bases at least Q20  
Insert size: 156000; agarose-fp  
Insert size: 162688; sum-of-contigs  
Quality coverage: 4.08 in Q20 bases; sum-of-contigs  
Quality coverage: 3.98 in Q20 bases; sum-of-contigs  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 24 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\*  
1  
\* 3018: contig of 3018 bp in length  
\* 3019 3118: gap of unknown length  
\* 3119 6167: contig of 3049 bp in length  
\* 6168 6267: gap of unknown length  
\* 6268 9294: contig of 3027 bp in length  
\* 9295 9395: gap of unknown length  
\* 9395 12600: contig of 3206 bp in length  
\* 12601 12700: gap of unknown length  
\* 12701 16144: contig of 3444 bp in length  
\* 16145 16244: gap of unknown length  
\* 16245 20910: contig of 4666 bp in length  
\* 20911 21010: gap of unknown length  
\* 21011 25128: contig of 4119 bp in length  
\* 25130 25229: gap of unknown length  
\* 25230 30373: contig of 5144 bp in length  
\* 30374 30474: gap of unknown length  
\* 30475 34806: contig of 4333 bp in length  
\* 34807 34907: gap of unknown length  
\* 34908 39962: contig of 5056 bp in length  
\* 39963 40062: gap of unknown length  
\* 40063 46152: contig of 6090 bp in length  
\* 46153 46252: gap of unknown length  
\* 46253 52245: contig of 5993 bp in length  
\* 52246 52345: gap of unknown length  
\* 52346 58622: contig of 6277 bp in length  
\* 58623 58722: gap of unknown length  
\* 58723 58105: contig of 6383 bp in length  
\* 58106 65205: gap of unknown length  
\* 65206 72475: contig of 7270 bp in length  
\* 72476 72575: gap of unknown length  
\* 72576 80000: contig of 7425 bp in length  
\* 80001 80100: gap of unknown length  
\* 80101 90367: contig of 10267 bp in length  
\* 90368 90467: gap of unknown length

\* 90468 99743: contig of 9276 bp in length  
\* 99744 99843: gap of unknown length  
\* 99844 114077: contig of 14234 bp in length  
\* 114078 114177: gap of unknown length  
\* 114178 128059: contig of 13882 bp in length  
\* 128060 128159: gap of unknown length  
\* 128160 144663: contig of 16504 bp in length  
\* 144664 144763: gap of unknown length  
\* 144764 160610: contig of 15847 bp in length  
\* 160611 160710: gap of unknown length  
\* 160711 162735: contig of 2025 bp in length  
\* 162736 162835: gap of unknown length  
\* 162836 164988: contig of 2153 bp in length.  
Location/Qualifiers  
1. 164988  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="4"  
/clone="RP11-162D2"  
1. 3018  
/note="assembly\_name:Contig10"  
3119. 6167  
/note="assembly\_name:Contig11"  
6268. 9294  
/note="assembly\_name:Contig12"  
9395. 12600  
/note="assembly\_name:Contig13"  
12701. 16144  
/note="assembly\_name:Contig14  
clone\_end:17  
vector\_side:right"  
16245. 20910  
/note="assembly\_name:Contig15"  
21011. 25129  
/note="assembly\_name:Contig16"  
25230. 30373  
/note="assembly\_name:Contig17"  
30474. 34806  
/note="assembly\_name:Contig18"  
34907. 39962  
/note="assembly\_name:Contig19"  
40063. 46152  
/note="assembly\_name:Contig20"  
46253. 52245  
/note="assembly\_name:Contig21"  
52346. 58622  
/note="assembly\_name:Contig22"  
58723. 65105  
/note="assembly\_name:Contig23"  
65206. 72475  
/note="assembly\_name:Contig24"  
72576. 80000  
/note="assembly\_name:Contig25"  
80101. 90367  
/note="assembly\_name:Contig26"  
90468. 99743  
/note="assembly\_name:Contig27"  
99844. 114077  
/note="assembly\_name:Contig28"  
114178. 128059  
/note="assembly\_name:Contig29"  
128160. 144663  
/note="assembly\_name:Contig30"  
144764. 160610  
/note="assembly\_name:Contig31"  
160711. 162735  
/note="assembly\_name:Contig8"  
162836. 164988  
/note="assembly\_name:Contig9"

BASE COUNT 48065 a 33578 c 34317 g 46717 t 2311 others  
ORIGIN  
Query Match 95.2%; Score 94.2; DB 2; Length 164988;

Best Local Similarity 97.0%; Pred. No. 5.4e-22;  
Matches 96; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

1 TCACCCCTGAGCAGCAATGTTGGCATCTGTAAAGACCACTAGATCCAGCA 60  
Db 70146 TCACCCCTGAGCAGCAATGTTGGCATCTGTAAAGACCACTAGATCCAGCA 70205

61 GCCCAGACCCCTTCTTTGTGTGTCAGAAAGCCGGGAAA 99  
Db 70206 GCCCAGACCCCTTCTTTGTGTGTCAGAAAGCAGGAAA 70244

RESULT 10  
AC073066 169286 bp DNA linear PRI 09-JAN-2002  
LOCUS Homo sapiens BAC clone RP11-149P12 from 2, complete sequence.  
DEFINITION AC073066  
AC073066  
AC073066.5 GI:15778739  
KEYWORDS HTG.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE  
AUTHORS 1 (bases 1 to 169286)  
TITLE Sultun,J.E. and Waterston,R.  
JOURNAL Toward a complete human genome sequence  
MEDLINE Genome Res. 8 (11), 1097-1108 (1998)  
99063792  
PUBMED 9847074

REFERENCE  
AUTHORS 2 (bases 1 to 169286)  
TITLE Desai,A., Cordum,H., Maupin,R. and Elliott,G.  
JOURNAL The sequence of Homo sapiens BAC clone RP11-149P12  
REFERENCE 3 (bases 1 to 169286)  
AUTHORS Waterston,R.H.  
TITLE Direct Submission  
JOURNAL Submitted (08-JUN-2000) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA

REFERENCE  
AUTHORS 4 (bases 1 to 169286)  
TITLE Waterston,R.H.  
JOURNAL Direct Submission  
Submitted (26-SEP-2001) Genome Sequencing Center, Washington  
University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA

REFERENCE  
AUTHORS 5 (bases 1 to 169286)  
TITLE Waterston,R.  
JOURNAL Direct Submission  
Submitted (09-JAN-2002) Department of Genetics, Washington  
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
On Sep 26, 2001 this sequence version replaced gi:15528927.

Center: Washington University Genome Sequencing Center  
Center code: MUGSC  
Web site: <http://genome.wustl.edu/gsc>  
Contact: [sapiens@wustl.edu](mailto:sapiens@wustl.edu)  
Summary Statistics  
Center project name: H\_NH0149P12

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality > 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:  
Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:  
The RPc1-11 human BAC library was made from the blood of one male donor, as described by Osoegawa,K., Moon,P.Y., Zhao,B., Frengen,E., Tatenoe,M., Catanesse,J.J. and de Jong,P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)  
VECTOR: PBAC3.6

NEIGHBORING SEQUENCE INFORMATION:  
Actual start of this clone is at base position 1 of RP11-149P12; actual end is at base position 169286 of RP11-149P12.

Data from AC024424 was used to finish this clone, AC073066.  
Polymorphisms have been identified between AC073066 and AC024424.

#### FEATURES

##### source

1. 169286  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="2"  
/map="2"  
/clone="RP11-149P12"  
/clone\_lib="RPc1-11"  
15. 103  
/rpt\_family="L1"  
117. 421  
/rpt\_family="L1"  
658. 690  
/rpt\_family="AT-rich"  
1275. 1321  
/rpt\_family="AT-rich"  
1401. 1786  
/rpt\_family="MaLR"  
2093. 2122  
/rpt\_family="(TTTTC)n"  
2160. 2192  
/rpt\_family="AT-rich"  
2174. 2293  
/rpt\_family="L1"  
2832. 2893  
/rpt\_family="AT-rich"  
3235. 3373  
/rpt\_family="L2"  
3597. 3850  
/rpt\_family="Alu"  
3630. 3662  
/rpt\_family="(A)n"  
3818. 3879  
/rpt\_family="A-rich"  
4072. 4489  
/rpt\_family="MER2\_type"  
5015. 5393  
/rpt\_family="MaLR"  
5520. 5640  
/rpt\_family="L2"  
5700. 6149  
/rpt\_family="ERV1"  
6185. 6525  
/rpt\_family="ERV1"  
6526. 7677  
/rpt\_family="L1"  
7662. 7693  
/rpt\_family="AT-rich"  
7678. 7792  
/rpt\_family="ERV1"  
7878. 8058  
repeat\_region



```
repeat_region      /rpt_family="ERV1"
                    8162..8483
repeat_region      /rpt_family="ERV1"
                    8491..8791
repeat_region      /rpt_family="Alu"
                    8937..9063
repeat_region      /rpt_family="Alu"
                    9067..9669
repeat_region      /rpt_family="ERV1"
                    9682..9625
repeat_region      /rpt_family="ERV1"
                    9871..10540
repeat_region      /rpt_family="L1"
                    10546..10733
repeat_region      /rpt_family="L1"
                    10734..11190
repeat_region      /rpt_family="L1"
                    15973..16198
repeat_region      /rpt_family="ERV1"
                    17703..17734
repeat_region      /rpt_family="(Tg)n"
                    1777..18013
repeat_region      /rpt_family="L2"
                    18844..18871
repeat_region      /rpt_family="AT_rich"
                    18989..19088
repeat_region      /rpt_family="(TA)n"
                    19113..19140
repeat_region      /rpt_family="(TATATG)n"
                    19350..19537
repeat_region      /rpt_family="MALR"
                    19553..19710
repeat_region      /rpt_family="L1"
                    19748..19770
repeat_region      /rpt_family="AT_rich"
                    19940..20478
repeat_region      /rpt_family="ERV1"
                    20453..20495
repeat_region      /rpt_family="(TCTA)n"
                    20533..20930
repeat_region      /rpt_family="L1"
                    20643..20688
repeat_region      /rpt_family="AT_rich"
                    20938..21055
repeat_region      /rpt_family="L1"
                    22130..22220
repeat_region      /rpt_family="L2"
                    22440..22559
repeat_region      /rpt_family="L1"
                    22677..23841
repeat_region      /rpt_family="MER2_type"
                    23210..23257
repeat_region      /rpt_family="AT_rich"
                    23351..23432
repeat_region      /rpt_family="AT_rich"
                    23846..24390
repeat_region      /rpt_family="ERV1"
                    24466..24606
repeat_region      /rpt_family="L1"
                    24607..24756
repeat_region      /rpt_family="Alu"
                    24758..25541
repeat_region      /rpt_family="ERV1"
                    25600..26211
repeat_region      /rpt_family="L1"
                    26212..26607
repeat_region      /rpt_family="ERV1"
                    26608..27012

Query Match      95.28; Score 94.2; DB 9; Length 169286;
Best Local Similarity 97.08; Pred. No. 5,4e-22;
Matches 96; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

```
QY      1 TCACCCGAGCGCAAGGCAATGTGGCATGTGCTGAAGACCACTGAATCCAGCA 60
          |||
Db      83577 TCACCCGAGCGCAAGGCAATGTGGCATGTGCTGAAGACCACTGAATCCAGCA 83518
          |||

QY      61 GCCCAGACCCCTTCTTGTGTCAGAAAGCGCGGAAA 99
          |||
Db      83517 GCCCAGACCCCTTCTTGTGTCAGAAAGCGCGGAAA 83479
          |||

RESULT 11
AC024033/c      169462 bp DNA linear HTG 24 JAN-2002
LOCUS      Homo sapiens chromosome 3 clone RP11-215L17, WORKING DRAFT
DEFINITION      SEQUENCE, 18 unordered pieces.
ACCESSION      AC024033.5 GI:18308831
VERSION      AC024033.5
KEYWORDS      HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens.
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE      1 (bases 1 to 169462)
AUTHORS      Waterston,R.H.
TITLE      The sequence of Homo sapiens clone
REFERENCE      2 (bases 1 to 169462)
AUTHORS      Waterston,R.H.
TITLE      Direct Submission
COMMENT      Submitted (20-FEB-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
            On Jan 24, 2002 this sequence version replaced gi:8572509.

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc/index.shtml
Contact: submissions@watson.wustl.edu
Project Information
Center project name: H_NH0215L17
----- Summary Statistics -----
Sequencing vector: M13; 89%
Sequencing vector: plasmid; 11%
Chemistry: Dye-primer ET; 89% of reads
Chemistry: Dye-terminator Big Dye; 11% of reads
Assembly program: Phrap; version 0.990319
Consensus quality: 162787 bases at least Q40
Consensus quality: 165310 bases at least Q30
Consensus quality: 166295 bases at least Q20
Insert size: 146000; agarose-fp
Insert size: 169084; sum-of-contigs
Quality coverage: 4.10 in Q20 bases; agarose-fp
Quality coverage: 4.67 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 18 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1      1088: contig of 1088 bp in length
*      1089      1188: gap of unknown length
*      1189      2342: contig of 1154 bp in length
*      2343      2442: gap of unknown length
*      2443      4017: contig of 1575 bp in length
*      4018      4117: gap of unknown length
*      4118      5126: contig of 1009 bp in length
*      5127      5226: gap of unknown length
*      5227      8869: contig of 3643 bp in length
*      8870      8969: gap of unknown length
```



## restriction digest.

MAPPING INFORMATION: Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

## SOURCE INFORMATION:

The RPT-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Firengen, E., Tatenio, M., Catanesi, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://pacpac.med.buffalo.edu>)

## VECTOR: pBACe3.6

## NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP11-188P17. Actual start of this clone is at base position 1 of RP11-335L23; actual end is at base position 177720 of RP11-335L23.

Data from AC080070 and AC079766 was used to finish this clone, AC019193. Polymorphisms have been identified between AC019193 and AC080070.

## FEATURES

## source

1.177720 location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="4"  
/map="4"  
/clone="RP11-335L23"  
/clone\_id="RPT-11"  
1050..1350  
/rpt\_family="Alu"  
2334..2353  
/rpt\_family="(CA)n"  
4076..4120  
/rpt\_family="Alu"  
4121..5207  
/rpt\_family="L1"  
5085..5089  
/note="similar to Homo sapiens EST BG994784 (NID:914398654)"  
5204..5835  
/rpt\_family="L1"  
6169..6237  
/rpt\_family="L1"  
6238..6546  
/rpt\_family="Alu"  
6516..6550  
/rpt\_family="A-rich"  
6547..6737  
/rpt\_family="L1"  
6738..7036  
/rpt\_family="Alu"  
6860..6881  
/rpt\_family="AT-rich"  
7037..7112  
/rpt\_family="L1"  
7163..8060  
/rpt\_family="L1"  
8063..8444  
/rpt\_family="MALR"  
8506..8814  
/rpt\_family="Alu"  
8788..8814  
/rpt\_family="AT-rich"  
8972..9283  
/rpt\_family="Alu"  
9254..9283  
/rpt\_family="(A)n"

repeat\_region 9284..9306  
/rpt\_family="(TA)n"  
repeat\_region 9306..9334  
/rpt\_family="(CA)n"  
repeat\_region 9403..9557  
/rpt\_family="L1"  
repeat\_region 9555..9666  
/rpt\_family="L1"  
repeat\_region 9672..10122  
/rpt\_family="L1"  
repeat\_region 9831..9902  
/rpt\_family="L1"  
repeat\_region 10150..10275  
/rpt\_family="GA-rich"  
repeat\_region 10251..10433  
/rpt\_family="L1"  
repeat\_region 10288..10603  
/rpt\_family="L1"  
repeat\_region 10608..10660  
/rpt\_family="L1"  
repeat\_region 10859..11020  
/rpt\_family="MERL-type"  
repeat\_region 11153..11521  
/rpt\_family="L1"  
repeat\_region 12033..12483  
/rpt\_family="MALR"  
misc\_feature /note="match to EST AA309275 (NID:91961665)"  
repeat\_region 1281..12581  
/rpt\_family="Alu"  
repeat\_region 12600..12624  
/rpt\_family="(T)n"  
repeat\_region 13032..13102  
/rpt\_family="MIR"  
repeat\_region 13105..13540  
/rpt\_family="L1"  
repeat\_region 13522..13560  
/rpt\_family="L1"  
repeat\_region 14010..14262  
/rpt\_family="AT-rich"  
repeat\_region 14249..15274  
/rpt\_family="L1"  
repeat\_region 15275..15494  
/rpt\_family="MERL-type"  
repeat\_region 15495..15633  
/rpt\_family="L1"  
repeat\_region 15836..15904  
/rpt\_family="MER2-type"  
repeat\_region 15895..16327  
/rpt\_family="MER2-type"  
repeat\_region 16329..16885  
/rpt\_family="L1"  
repeat\_region 16886..17181  
/rpt\_family="Alu"  
repeat\_region 17182..17455  
/rpt\_family="L1"  
repeat\_region 17456..17763  
/rpt\_family="Alu"  
repeat\_region 17741..17773  
/rpt\_family="(CAA)n"  
repeat\_region 17764..17827  
/rpt\_family="L1"  
repeat\_region 17828..18120  
/rpt\_family="Alu"  
repeat\_region 18121..18805  
/rpt\_family="L1"  
repeat\_region 18792..20289  
/rpt\_family="L1"  
repeat\_region 24661..24699  
/rpt\_family="BRV1"  
repeat\_region 24684..24981  
/rpt\_family="BRV1"

Query Match 95.2% Score 94.2; DB 9; Length 177720;  
Best Local Similarity 97.0%; Pred. No. 5.4e-22;

Matches 96; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TCACCCCTGAGCAAGCATGTGGCATCTGTAAAGACCACTAGATCCAGCA 60  
|||||  
Db 20339 TCACCCCTGAGCAAGCAATGTGGCATCTGTAAAGACCACTAGATCCAGCA 20398  
|||||

QY 61 GCCCAGACCCCTTTCTTTGTGTCAAGAAAGCGCGGAA 99  
|||||  
Db 20399 GCCCAGACCCCTTTCTTTGTGTCAAGAAAGCGGAA 20437  
|||||

RESULT 13  
AC027752 186723 bp DNA linear HTG 04-MAY-2001  
LOCUS Homo sapiens chromosome 11 clone RP11-400F23 map 11, WORKING DRAFT  
DEFINITION  
SEQUENCE 50 unordered pieces.  
AC027752  
VERSION AC027752.2 GI:8671973  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 186723)  
AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
TITLE Homo sapiens chromosome 11, clone RP11-400F23  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 186723)  
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,  
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,  
Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G.,  
Campopiano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,  
Collamore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,  
Dodge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,  
Galagan,J., Gardyna,S., Ginde,S., Govette,M., Graham,L.,  
Grand-Pierre,N., Grant,G., Hages,B., Heaford,A., Horton,L.,  
Howard,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,  
Klein,J., Lacroque,K., Lamazares,R., Landers,T., Lehoczy,J.,  
Levine,R., Lieu,C., Liu,G., Locke,K., MacDonald,P., Marquis,N.,  
McCarthy,M., McEwan,P., McGuirk,A., McKernan,K., McSheeters,R.,  
Meldrum,J., Menues,L., Mihova,T., Miranda,C., Mienga,V., Morrow,J.,  
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,  
O'Neill,D., Olivari,T.M., Oliver,J., Peterson,K., Plettre,N.,  
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,  
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,  
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,  
Thesler,S., Theodore,J., Tlreil,A., Travers,M., Triggillo,J.,  
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,  
Young,G., Zainoun,J., Zimmer,A. and Zody,M.  
Direct Submission  
Submitted (01-APR-2000) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Jun 23, 2000 this sequence version replaced gi:7382576.  
ALL repeats were identified using RepeatMasker:  
Smit,A.F.A. & Green,P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: http://www-seq.wi.mit.edu  
----- Project Information  
Center project name: L8231  
Center clone name: 400\_F\_23  
----- Summary Statistics  
Sequencing vector: M13, M7815, 100% of reads  
Chemistry: Dye-terminator Big Dye, 100% of reads  
Assembly program: Phrap, version 0.960731  
Consensus quality: 161774 bases at least Q40  
Consensus quality: 173958 bases at least Q30  
Consensus quality: 178770 bases at least Q20  
Insert size: 182000; agarose-fp  
Insert size: 181823; sum-of-ctrls

Quality coverage: 3.2 In Q20 bases; agarose-fp  
Quality coverage: 3.2 In Q20 bases; sum-of-ctrls  
-----  
NOTE: This is a 'working draft' sequence. It currently  
consists of 50 contigs. The true order of the pieces  
is not known and their order in this sequence record is  
arbitrary. Gaps between the contigs are represented as  
runs of N, but the exact sizes of the gaps are unknown.  
This record will be updated with the finished sequence  
as soon as it is available and the accession number will  
be preserved.  
1 1069: contig of 1069 bp in length  
1070 1169: gap of 100 bp  
1170 2293: contig of 1124 bp in length  
2294 2393: gap of 100 bp  
2394 3428: contig of 1035 bp in length  
3429 3528: gap of 100 bp  
3529 4652: contig of 1124 bp in length  
4653 4752: gap of 100 bp  
4753 6555: contig of 1803 bp in length  
6556 6655: gap of 100 bp  
6656 8503: contig of 1848 bp in length  
8504 8603: gap of 100 bp  
8604 10200: contig of 1597 bp in length  
10201 10300: gap of 100 bp  
10301 11338: contig of 1038 bp in length  
11339 11438: gap of 100 bp  
11439 13118: contig of 1680 bp in length  
13119 13218: gap of 100 bp  
13219 14467: contig of 1249 bp in length  
14468 14567: gap of 100 bp  
14568 15606: contig of 1039 bp in length  
15607 15706: gap of 100 bp  
15707 17281: contig of 1575 bp in length  
17282 17381: gap of 100 bp  
17382 18795: contig of 1414 bp in length  
18796 18895: gap of 100 bp  
18896 20660: contig of 1765 bp in length  
20661 20760: gap of 100 bp  
20761 22420: contig of 1660 bp in length  
22421 22520: gap of 100 bp  
22521 24824: contig of 2304 bp in length  
24825 24924: gap of 100 bp  
24925 26647: contig of 1733 bp in length  
26648 26747: gap of 100 bp  
26748 28100: contig of 1333 bp in length  
28101 28200: gap of 100 bp  
28201 29210: contig of 1010 bp in length  
29211 29310: gap of 100 bp  
29311 31398: contig of 2088 bp in length  
31399 31498: gap of 100 bp  
31499 33852: contig of 2354 bp in length  
33853 33952: gap of 100 bp  
33953 35503: contig of 1551 bp in length  
35504 35603: gap of 100 bp  
35604 38508: contig of 2905 bp in length  
38509 38608: gap of 100 bp  
38609 41035: contig of 2427 bp in length  
41036 41135: gap of 100 bp  
41136 44494: contig of 3359 bp in length  
44495 44594: gap of 100 bp  
44595 47972: contig of 3378 bp in length  
47973 48072: gap of 100 bp  
48073 51596: contig of 3524 bp in length  
51597 51696: gap of 100 bp  
51697 55363: contig of 3667 bp in length  
55364 55463: gap of 100 bp  
55464 58776: contig of 3213 bp in length  
58777 62286: contig of 3510 bp in length  
62287 62386: gap of 100 bp  
62387 65381: contig of 2995 bp in length  
65382 65481: gap of 100 bp

29 TCACCCCTGAGCGCAAGGCAATGTTGGCATGCTGGTAAGGACCACCTAGAATCCAGCA 60

```
FEATURES
SOURCE
Location/Qualifiers
1..340000
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="21"
/map="21q22.3"
<1..125946
/organism="Homo sapiens"
```



0Y 61 GCCCAGACCCCTTCTTGTGTCAGAAAGCGCGGAAA 99  
|||||  
DB 247087 GCCCGACCCCTTCTTGTGTCAGAAAGCGCGGAAA 247049

RESULT 15  
AC093393  
LOCUS AC093393 120450 bp DNA linear PRI 09-JAN-2002  
DEFINITION Homo sapiens BAC clone RP11-816B18 from 2, complete sequence.  
ACCESSION AC093393  
VERSION AC093393.3 GI:18042521  
KEYWORDS HTG.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE  
AUTHORS 1 (bases 1 to 120450)  
TITLE Sulston, J.E. and Waterston, R.  
JOURNAL Toward a complete human genome sequence  
MEDLINE Genome Res. 8 (11), 1097-1108 (1998)  
PMID 99063792

REFERENCE  
AUTHORS 2 (bases 1 to 120450)  
TITLE Harris, A. and Abbott, A.  
JOURNAL The sequence of Homo sapiens BAC clone RP11-816B18  
UNPUBLISHED (2001)  
AUTHORS 3 (bases 1 to 120450)  
JOURNAL Waterston, R.H.  
UNPUBLISHED (2001)

REFERENCE  
AUTHORS Direct Submission  
JOURNAL Submitted (21-AUG-2001) Genome Sequencing Center, Washington  
UNIVERSITY School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
4 (bases 1 to 120450)  
JOURNAL Waterston, R.H.  
UNPUBLISHED (2001)

REFERENCE  
AUTHORS Direct Submission  
JOURNAL Submitted (03-JAN-2002) Genome Sequencing Center, Washington  
UNIVERSITY School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
5 (bases 1 to 120450)  
JOURNAL Waterston, R.H.  
UNPUBLISHED (2001)

REFERENCE  
AUTHORS Direct Submission  
JOURNAL Submitted (09-JAN-2002) Department of Genetics, Washington  
UNIVERSITY, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
On Jan 3, 2002 this sequence version replaced g1.16874923.  
COMMENT ----- Genome Center  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: <http://genome.wustl.edu/gsc>  
Contact: [sapliens@wustl.edu](mailto:sapliens@wustl.edu)  
----- Summary Statistics  
Center project name: H\_NH0816B18

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality > 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:  
Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis MO. For additional information about the map position of this sequence, see <http://genome.wustl.edu/gsc>

SOURCE INFORMATION:

#### FEATURES

source

The RPCT-11 human BAC library was made from the blood of one male donor, as described by Osoegawa, K., Moon, P.Y., Zhao, B., Frengen, E., Tatenno, M., Catanese, J.J. and de Jong, P.J. (1996) An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pletier de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)  
VECTOR: pBAC3.6  
NEIGHBORING SEQUENCE INFORMATION:  
The clone sequenced to the left is RP11-330A12; the clone sequenced to the right is RP11-332N21, 2000 bp overlap. Actual start of this clone is at base position 1 of RP11-816B18; actual end is at base position 64829 of RP11-332N21.

Data from AC017050 was used to finish AC093393. Polymorphisms have been identified between AC017050, AC069303 and AC093393. There is single plasmid coverage from 11435 to 114413.

Location/Qualifiers

1..120450

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="2"

/map="2"

/clone\_1id="RP11-816B18"

/clone\_1id="RPCT-11"

1..79

/rpt\_family="Alu"

82..143

/rpt\_family="MALR"

929..1108

/rpt\_family="L2"

1112..1374

/rpt\_family="MALR"

1291..1292

/note="similar to Homo sapiens EST AL537709  
(NID:g12801202)"

complement(1344..1345)

/note="similar to Homo sapiens EST BF215435  
(NID:g11109021)"

complement(1344..1345)

/note="similar to Homo sapiens EST BG502858  
(NID:g13464375)"

complement(1344..1345)

/note="similar to Homo sapiens EST BG506376  
(NID:g13467893)"

complement(1344..1345)

/note="similar to Homo sapiens EST BG532982  
(NID:g13524521)"

1377..2068

/note="similar to Homo sapiens EST A0117318  
(NID:g10932276)"

1384..2066

/note="similar to Homo sapiens EST B1549051  
(NID:g15436363)"

1385..1693

/note="similar to Homo sapiens EST BG506376  
(NID:g13467893)"

1385..1691

/note="similar to Homo sapiens EST BF215435  
(NID:g11109021)"

1385..1690

/note="similar to Homo sapiens EST BG532982  
(NID:g13524521)"

1385..1480

/note="similar to Homo sapiens EST BG502858  
(NID:g13464375)"

1395..2350

/note="similar to EST BM047694 (NID:g16776961)"

1395..1767

/note="similar to EST BM044716 (NID:g16773983)"

1397..2113

/note="similar to Homo sapiens EST BF786425

misc-feature

misc-feature

misc-feature

misc-feature

misc-feature

misc-feature

misc-feature

misc-feature

misc-feature

misc-feature

misc-feature

misc-feature

misc-feature

misc-feature

misc-feature

misc-feature

misc-feature

misc-feature

misc-feature

misc-feature

misc-feature

misc-feature

misc-feature

Search completed: May 2, 2003, 12:52:49  
Job time : 826.993 secs

```
misc_feature (NTD:g10207623)"
1398..1668 /note="similar to Homo sapiens EST BF678628
(NID:g11952523)"
misc_feature 1401..2046 /note="similar to Homo sapiens EST BG257033
(NID:g12766849)"
misc_feature 1401..1576 /note="similar to Homo sapiens EST AW954663
(NID:g8144346)"
misc_feature 1403..1867 /note="similar to Homo sapiens EST AA309125
(NID:g1961450)"
misc_feature 1404..1876 /note="similar to Homo sapiens EST BE960674
(NID:g10571483)"
misc_feature 1404..1693 /note="similar to Homo sapiens EST BF028041
(NID:g10735662)"
misc_feature 1405..2220 /note="similar to Homo sapiens EST BG677927
(NID:g13909324)"
misc_feature 1405..1974 /note="similar to Homo sapiens EST BG292090
(NID:g13050544)"
misc_feature 1410..2273 /note="similar to Homo sapiens EST AL537709
(NID:g12801202)"
misc_feature 1410..1794 /note="similar to Homo sapiens EST BG006478
(NID:g12449699)"
misc_feature 1414..2462 /note="similar to Homo sapiens EST AL513628
(NID:g12777122)"
misc_feature 1416..2116 /note="similar to Homo sapiens EST BM012067
(NID:g16526421)"
misc_feature 1420..2125 /note="similar to Homo sapiens EST AL515803
(NID:g12779296)"
misc_feature 1437..1915 /note="similar to Homo sapiens EST BI258319
(NID:g14814547)"
misc_feature 1437..1911 /note="similar to Homo sapiens EST BE295078
(NID:g9178623)"
misc_feature 1437..1759 /note="similar to Homo sapiens EST BI114149
(NID:g14565050)"
misc_feature 1437..1661 /note="similar to Homo sapiens EST AA300977
(NID:g1953330)"
misc_feature 1443..2034 /note="similar to Homo sapiens EST BG035997
(NID:g12430691)"
misc_feature 1446..2158 /note="similar to Homo sapiens EST BG618127
(NID:g13669498)"
misc_feature 1510..2042 /note="similar to Homo sapiens EST BG502858
(NID:g13464375)"
```

Query Match 94.1%; Score 93.2; DB 9; Length 120450;  
Best Local Similarity 96.9%; Pred. No. 1.2e-21;  
Matches 95; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
QY 1 TCACCCCTGAGCGCAAGCAATGTTGGCATCTGTAAAGACCACTGAATCCAGCA 60
DB 24603 TCACCCCTGAGCGCAAGCAATGTTGGCATCTGTAAAGACCACTGAATCCAGCA 24662
QY 61 GCCCAGACCCCTTCTTTGTGTCAGAAAGGGGGA 98
DB 24663 GCCCAGACCCCTTCTTTGTGTCAGAAAGGGGGA 24700
```



GenCore version 5.1.5  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 2, 2003, 11:21:20 ; Search time 157.717 Seconds  
(without alignments)  
1413.592 Million cell updates/sec

Title: US-09-719-554-3\_COPY\_2502\_2600

Perfect score: 99  
Sequence: 1 tcacccttgagcgcaaaagc.....tggtcaagaagcgaggaa 99

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

N.GeneSeq\_101002:\*

- 1: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:\*
- 2: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:\*
- 3: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:\*
- 4: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:\*
- 5: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:\*
- 6: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:\*
- 7: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:\*
- 8: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:\*
- 9: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:\*
- 10: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:\*
- 11: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:\*
- 12: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:\*
- 13: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:\*
- 14: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:\*
- 15: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:\*
- 16: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:\*
- 17: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:\*
- 18: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:\*
- 19: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:\*
- 20: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:\*
- 21: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:\*
- 22: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:\*
- 23: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:\*
- 24: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	99	100.0	575	23	AA572226
2	99	100.0	2049	23	AA590991
3	99	100.0	4805	23	AA573604
4	99	100.0	4805	23	AA573910
5	99	100.0	4805	23	AA576462
6	99	100.0	4805	23	AA584191
7	99	100.0	6507	23	AA573607
8	99	100.0	6507	23	AA573914
9	99	100.0	6507	23	AA592665

10	99	100.0	8279	23	AA576474	DNA encoding novel
11	99	100.0	8294	21	AA584209	DNA encoding novel
12	99	100.0	10495	21	AA597929	Human retroviral S
13	99	100.0	56093	24	ABL61744	Colon adenocarcino
14	97.4	98.4	1687	23	AA572234	DNA encoding novel
15	97.4	98.4	1729	23	AA577018	DNA encoding novel
16	94.2	95.2	1493	23	AA567606	DNA encoding novel
17	94.2	95.2	3831	23	AA571727	DNA encoding novel
18	92.6	93.5	865	23	AA572232	DNA encoding novel
19	92.6	93.5	1307	23	AA571726	DNA encoding novel
20	92.6	93.5	46340	21	ABN97978	Human retroviral S
21	91	91.9	1799	23	AA577017	DNA encoding novel
22	91	91.9	23855	22	AA199411	Human excretory re
23	91	91.9	23855	22	AA163761	Human kidney relat
24	89.4	90.3	510	23	AA576203	DNA encoding novel
25	89.4	90.3	510	23	AA577307	DNA encoding novel
26	89.4	90.3	510	23	AA591839	DNA encoding novel
27	89.4	90.3	1081	23	AA577016	DNA encoding novel
28	87.8	88.7	309	23	AA592668	DNA encoding novel
29	87.8	88.7	435	23	AA592494	DNA encoding novel
30	87.8	88.7	435	23	AA592664	DNA encoding novel
31	87.8	88.7	852	23	AA573694	DNA encoding novel
32	87.8	88.7	954	23	AA572223	DNA encoding novel
33	87.8	88.7	954	23	AA590990	DNA encoding novel
34	87.8	88.7	1539	23	AA573691	DNA encoding novel
35	87.8	88.7	3205	23	AA572236	DNA encoding novel
36	86.2	87.1	1719	23	AA581560	DNA encoding novel
37	85.8	86.7	924	23	AA572224	DNA encoding novel
38	85.8	86.7	1846	23	AA573605	DNA encoding novel
39	85.8	86.7	1846	23	AA584192	DNA encoding novel
40	85.2	86.1	804	23	AA572229	DNA encoding novel
41	85.2	86.1	867	23	AA576195	DNA encoding novel
42	85.2	86.1	1249	23	AA573698	DNA encoding novel
43	75.8	76.6	1518	22	AA564941	DNA encoding novel
44	75.8	76.6	3331	22	AA522442	Human cDNA encodin
45	75.8	76.6	3690	22	AA522678	Human cDNA encodin

#### ALIGNMENTS

RESULT 1	AA572226	standard; cDNA; 575 BP.
ID	AA572226;	
AC	AA572226;	
XX		
DT	13-FEB-2002 (first entry)	
XX		
DE	DNA encoding novel human diagnostic protein #8030.	
XX		
KW	Human; chromosome mapping; gene mapping; gene therapy; forensic;	
KW	food supplement; medical imaging; diagnostic; genetic disorder; ss.	
XX		
OS	Homo sapiens.	
XX		
PN	WO200175067-A2.	
XX		
PD	11-OCT-2001.	
XX		
PF	30-MAR-2001; 2001MO-US08631.	
XX		
PR	31-MAR-2000; 2000US-0540217.	
XX		
PR	23-AUG-2000; 2000US-0649167.	
XX		
PA	(HYSE-) HYSEQ INC.	
XX		
PI	Dymanac RT, Liu C, Tang YT;	
XX		
DR	WPI; 2001-639362/73.	
XX		
PT	P-PSDB; ABG08039.	
XX		
PT	New isolated polynucleotide and encoded polypeptides, useful in	

PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
XX  
PS Claim 1; SEQ ID No 8030; 103pp; English.  
XX  
CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAs64197-AAS94564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 575 BP; 183 A; 150 C; 140 G; 102 T; 0 other;  
XX  
Query Match 100.0%; Score 99; DB 23; Length 575;  
Best Local Similarity 100.0%; Pred. No. 3.2e-26;  
Matches 99; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
XX  
OY 1 TCACCCCTGAGCGCAAGCAATGTGGCATGCTGTTAAAGCACCTAGATCAGCA 60  
Db 23 TCACCCCTGAGCGCAAGCAATGTGGCATGCTGTTAAAGCACCTAGATCAGCA 82  
OY 61 GCCCAGACCCCTTCTTTGTGTCAGAAAGCGCGGAAA 99  
Db 83 GCCCAGACCCCTTCTTTGTGTCAGAAAGCGCGGAAA 121  
XX  
RESULT 2  
AAS90991  
ID AAS90991 standard; cDNA; 2049 BP.  
XX  
AC AAS90991;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE DNA encoding novel human diagnostic protein #26795.  
XX  
KW Human: chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US08631.  
XX  
PR 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YF;  
XX  
DR WPI: 2001-639362/73.  
DR P-PSDB: ABG26804.

XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
XX  
PS Claim 1; SEQ ID No 26795; 103pp; English.  
XX  
CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (II) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAs64197-AAS94564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 2049 BP; 612 A; 532 C; 444 G; 461 T; 0 other;  
XX  
Query Match 100.0%; Score 99; DB 23; Length 2049;  
Best Local Similarity 100.0%; Pred. No. 5.1e-26;  
Matches 99; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
XX  
OY 1 TCACCCCTGAGCGCAAGCAATGTGGCATGCTGTTAAAGCACCTAGATCAGCA 60  
Db 23 TCACCCCTGAGCGCAAGCAATGTGGCATGCTGTTAAAGCACCTAGATCAGCA 82  
OY 61 GCCCAGACCCCTTCTTTGTGTCAGAAAGCGCGGAAA 99  
Db 83 GCCCAGACCCCTTCTTTGTGTCAGAAAGCGCGGAAA 121  
XX  
RESULT 3  
AAS73604  
ID AAS73604 standard; cDNA; 4805 BP.  
XX  
AC AAS73604;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE DNA encoding novel human diagnostic protein #9408.  
XX  
KW Human: chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US08631.  
XX  
PR 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YF;  
XX

DR WPI: 2001-639362/73.  
DR P-PSDB: ABG09417.

PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity

PS Claim 1; SEQ ID No 9408; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

SQ Sequence 4805 BP; 1400 A; 1173 C; 1009 G; 1223 T; 0 other;

Query Match 100.0%; Score 99; DB 23; Length 4805;  
Best Local Similarity 100.0%; Pred. No. 6.9e-26;  
Matches 99; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TCACCCCTGAGCGCAAGCAATGTTGGCATGCTGTAAAGACCAATCCAGCA 60  
|||||  
DB 2950 TCACCCCTGAGCGCAAGCAATGTTGGCATGCTGTAAAGACCAATCCAGCA 3009

OY 61 GCCCAGACCCCTTTCTTTGTGCTCAAGAAAGCGGGAAA 99  
|||||  
DB 3010 GCCCAGACCCCTTTCTTTGTGCTCAAGAAAGCGGGAAA 3048

RESULT 4  
AAS73910  
ID AAS73910 standard; cDNA; 4805 BP.

AC AAS73910;

DT 13-FEB-2002 (first entry)

DE DNA encoding novel human diagnostic protein #9714.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

OS Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.  
XX (HYSE-) HYSEQ INC.  
XX

PI Drmenac Rf, Liu C, Tang YT;  
XX  
DR WPI: 2001-639362/73.  
DR P-PSDB: ABG09723.

PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity

PS Claim 1; SEQ ID No 9714; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

SQ Sequence 4805 BP; 1400 A; 1173 C; 1009 G; 1223 T; 0 other;

Query Match 100.0%; Score 99; DB 23; Length 4805;  
Best Local Similarity 100.0%; Pred. No. 6.9e-26;  
Matches 99; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 TCACCCCTGAGCGCAAGCAATGTTGGCATGCTGTAAAGACCAATCCAGCA 60  
|||||  
DB 2950 TCACCCCTGAGCGCAAGCAATGTTGGCATGCTGTAAAGACCAATCCAGCA 3009

OY 61 GCCCAGACCCCTTTCTTTGTGCTCAAGAAAGCGGGAAA 99  
|||||  
DB 3010 GCCCAGACCCCTTTCTTTGTGCTCAAGAAAGCGGGAAA 3048

RESULT 5  
AAS76462

ID AAS76462 standard; cDNA; 4805 BP.

AC AAS76462;

DT 13-FEB-2002 (first entry)

DE DNA encoding novel human diagnostic protein #12266.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

OS Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX

PA (HYSE-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;  
XX  
XX WPI; 2001-639362/73.  
DR P-PSDB; ABG12275.  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity  
XX  
XX Claim 1; SEQ ID No 12266; 103bp; English.  
XX  
XX The invention relates to isolated polynucleotide (I) and  
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,  
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
XX and gene mapping, and in recombinant production of (II). The  
XX polynucleotides are also used in diagnostics as expressed sequence tags  
XX for identifying expressed genes. (I) is useful in gene therapy techniques  
XX to restore normal activity of (II) or to treat disease states involving  
XX (II). (II) is useful for generating antibodies against it, detecting or  
XX quantitating a polypeptide in tissue, as molecular weight markers and as  
XX a food supplement. (II) and its binding partners are useful in medical  
XX imaging of sites expressing (II). (I) and (II) are useful for treating  
XX disorders involving aberrant protein expression or biological activity.  
XX The polypeptide and polynucleotide sequences have applications in  
XX diagnostics, forensics, gene mapping, identification of mutations  
XX responsible for genetic disorders or other traits to assess biodiversity  
XX and to produce other types of data and products dependent on DNA and  
XX amino acid sequences. AAS64197-AAS94564 represent novel human  
XX diagnostic coding sequences of the invention.  
XX Note: The sequence data for this patent did not appear in the printed  
XX specification, but was obtained in electronic format directly from WIPO  
XX at ftp.wipo.int/pub/published\_pct\_sequences.  
SQ Sequence 4805 BP; 1400 A; 1173 C; 1009 G; 1223 T; 0 other;  
Query Match 100.0%; Score 99; DB 23; Length 4805;  
Best Local Similarity 100.0%; Pred. No. 6.9e-26;  
Matches 99; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 TCACCCCTGAGCGCAAGGCAATGTGGCATGCTGTAAAGCAGCACTGATTCAGCA 60  
|||||  
DB 2950 TCACCCCTGAGCGCAAGGCAATGTGGCATGCTGTAAAGCAGCACTGATTCAGCA 3009  
QY 61 GCCCAGACCCCTTCTTGTGTGTCAGAGAAAGCGGGAAA 99  
|||||  
DB 3010 GCCCAGACCCCTTCTTGTGTGTCAGAGAAAGCGGGAAA 3048  
RESULT 6  
AAS84191  
ID AAS84191 standard; cDNA; 4805 BP.  
XX  
AC AAS84191;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE DNA encoding novel human diagnostic protein #19995.  
XX  
KW Human: chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001MO-US08631.  
XX  
PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.  
XX  
XX (HYSE-) HYSEQ INC.  
PA  
XX  
PI Drmanac RT, Liu C, Tang YT;  
XX  
XX WPI; 2001-639362/73.  
DR P-PSDB; ABG20004.  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity  
XX  
XX Claim 1; SEQ ID No 19995; 103bp; English.  
XX  
XX The invention relates to isolated polynucleotide (I) and  
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,  
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
XX and gene mapping, and in recombinant production of (II). The  
XX polynucleotides are also used in diagnostics as expressed sequence tags  
XX for identifying expressed genes. (I) is useful in gene therapy techniques  
XX to restore normal activity of (II) or to treat disease states involving  
XX (II). (II) is useful for generating antibodies against it, detecting or  
XX quantitating a polypeptide in tissue, as molecular weight markers and as  
XX a food supplement. (II) and its binding partners are useful in medical  
XX imaging of sites expressing (II). (I) and (II) are useful for treating  
XX disorders involving aberrant protein expression or biological activity.  
XX The polypeptide and polynucleotide sequences have applications in  
XX diagnostics, forensics, gene mapping, identification of mutations  
XX responsible for genetic disorders or other traits to assess biodiversity  
XX and to produce other types of data and products dependent on DNA and  
XX amino acid sequences. AAS64197-AAS94564 represent novel human  
XX diagnostic coding sequences of the invention.  
XX Note: The sequence data for this patent did not appear in the printed  
XX specification, but was obtained in electronic format directly from WIPO  
XX at ftp.wipo.int/pub/published\_pct\_sequences.  
SQ Sequence 4805 BP; 1400 A; 1173 C; 1009 G; 1223 T; 0 other;  
Query Match 100.0%; Score 99; DB 23; Length 4805;  
Best Local Similarity 100.0%; Pred. No. 6.9e-26;  
Matches 99; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 TCACCCCTGAGCGCAAGGCAATGTGGCATGCTGTAAAGCAGCACTGATTCAGCA 60  
|||||  
DB 2950 TCACCCCTGAGCGCAAGGCAATGTGGCATGCTGTAAAGCAGCACTGATTCAGCA 3009  
QY 61 GCCCAGACCCCTTCTTGTGTGTCAGAGAAAGCGGGAAA 99  
|||||  
DB 3010 GCCCAGACCCCTTCTTGTGTGTCAGAGAAAGCGGGAAA 3048  
RESULT 7  
AAS73607  
ID AAS73607 standard; cDNA; 6507 BP.  
XX  
AC AAS73607;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE DNA encoding novel human diagnostic protein #9411.  
XX  
KW Human: chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001MO-US08631.

XX 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX (HYSE-) HXSEQ INC.  
XX Drmanac RT, Liu C, Tang YT;  
XX WPI: 2001-639362/73.  
DR P-PSDB: ABG09420.  
XX  
XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
PS Claim 1; SEQ ID No 9411; 103pp; English.  
XX  
XX The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 6507 BP; 1936 A; 1524 C; 1395 G; 1652 T; 0 other;  
Query Match 100.0%; Score 99; DB 23; Length 6507;  
Best Local Similarity 100.0%; Pred. No. 7.6e-26;  
Matches 99; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 TCACCCCTGAGCGCAAGGCAATGTTGGCATGCTGTAAGAGACCCTAGATCCAGCA 60  
|||||  
DB 4652 TCACCCCTGAGCGCAAGGCAATGTTGGCATGCTGTAAGAGACCCTAGATCCAGCA 4711  
OY 61 GCCCAGACCCCTTCTTGTGTGTCAGAAAGCGGGGAAA 99  
|||||  
DB 4712 GCCCAGACCCCTTCTTGTGTGTCAGAAAGCGGGGAAA 4750  
RESULT 8  
AAS73914  
ID AAS73914 standard; cDNA; 6507 BP.  
XX  
AC AAS73914;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE DNA encoding novel human diagnostic protein #9718.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US08631.  
PF 31-MAR-2000; 2000US-0540217.  
XX 23-AUG-2000; 2000US-0649167.  
PR 23-AUG-2000; 2000US-0649167.  
XX (HYSE-) HXSEQ INC.  
XX Drmanac RT, Liu C, Tang YT;  
XX WPI: 2001-639362/73.  
DR P-PSDB: ABG09727.  
XX  
XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
PS Claim 1; SEQ ID No 9718; 103pp; English.  
XX  
XX The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 6507 BP; 1936 A; 1524 C; 1395 G; 1652 T; 0 other;  
Query Match 100.0%; Score 99; DB 23; Length 6507;  
Best Local Similarity 100.0%; Pred. No. 7.6e-26;  
Matches 99; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 TCACCCCTGAGCGCAAGGCAATGTTGGCATGCTGTAAGAGACCCTAGATCCAGCA 60  
|||||  
DB 4652 TCACCCCTGAGCGCAAGGCAATGTTGGCATGCTGTAAGAGACCCTAGATCCAGCA 4711  
OY 61 GCCCAGACCCCTTCTTGTGTGTCAGAAAGCGGGGAAA 99  
|||||  
DB 4712 GCCCAGACCCCTTCTTGTGTGTCAGAAAGCGGGGAAA 4750  
RESULT 9  
AAS92665/c  
ID AAS92665 standard; cDNA; 6507 BP.  
XX  
AC AAS92665;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE DNA encoding novel human diagnostic protein #28469.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD

XX 11-OCT-2001.  
PD 30-MAR-2001: 2001MO-US08631.  
XX  
XX 31-MAR-2000: 2000US-0540217.  
PR 23-AUG-2000: 2000US-0649167.  
XX  
XX (HYSE-) HYSEQ INC.  
PA  
PI Drmanac RT, Liu C, Tang YF;  
XX WPI: 2001-639362/73.  
DR P-PSDB; ABG28478.  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity.  
XX  
XX Claim 1; SEQ ID No 28469; 103pp; English.  
XX  
CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_sequences.  
XX  
SQ Sequence 6507 BP; 1652 A; 1395 C; 1524 G; 1936 T; 0 other;  
XX  
Query Match 100.0%; Score 99; DB 23; Length 6507;  
Best Local Similarity 100.0%; Pred. No. 7.6e-26;  
Matches 99; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
XX  
QY 1 TCACCCCTGAGCGCAAGCAATGTTGGCATGCTGTAAGACCACTGAATCCAGCA 60  
DB 1856 TCACCCCTGAGCGCAAGCAATGTTGGCATGCTGTAAGACCACTGAATCCAGCA 1797  
QY 61 GCCCAGACCCCTTTCTTTGTGTCAGAAAGCGGGGAAA 99  
DB 1796 GCCCAGACCCCTTTCTTTGTGTCAGAAAGCGGGGAAA 1758  
XX  
RESULT 10  
AAS76474  
ID AAS76474 standard; cDNA; 8279 BP.  
XX  
AC AAS76474;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE DNA encoding novel human diagnostic protein #12278.  
XX  
KW Human: chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
OS Homo sapiens.

XX MO200175067-A2.  
PN  
XX 11-OCT-2001.  
PD 30-MAR-2001: 2001MO-US08631.  
XX  
XX 31-MAR-2000: 2000US-0540217.  
PR 23-AUG-2000: 2000US-0649167.  
XX  
XX (HYSE-) HYSEQ INC.  
PA  
PI Drmanac RT, Liu C, Tang YF;  
XX WPI: 2001-639362/73.  
DR P-PSDB; ABG12287.  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity.  
XX  
XX Claim 1; SEQ ID No 12278; 103pp; English.  
XX  
CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_sequences.  
XX  
SQ Sequence 8279 BP; 2474 A; 1992 C; 1770 G; 2043 T; 0 other;  
XX  
Query Match 100.0%; Score 99; DB 23; Length 8279;  
Best Local Similarity 100.0%; Pred. No. 8.3e-26;  
Matches 99; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
XX  
QY 1 TCACCCCTGAGCGCAAGCAATGTTGGCATGCTGTAAGACCACTGAATCCAGCA 60  
DB 4652 TCACCCCTGAGCGCAAGCAATGTTGGCATGCTGTAAGACCACTGAATCCAGCA 4711  
QY 61 GCCCAGACCCCTTTCTTTGTGTCAGAAAGCGGGGAAA 99  
DB 4712 GCCCAGACCCCTTTCTTTGTGTCAGAAAGCGGGGAAA 4750  
XX  
RESULT 11  
AAS84209  
ID AAS84209 standard; cDNA; 8294 BP.  
XX  
AC AAS84209;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE DNA encoding novel human diagnostic protein #20013.  
XX  
KW Human: chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

```
XX OS Homo sapiens.
XX PN WO200175067-A2.
XX PD 11-OCT-2001.
XX PF 30-MAR-2001; 2001WO-US08631.
XX PR 31-MAR-2000; 2000US-0540217.
XX PR 23-AUG-2000; 2000US-0649167.
XX PA (HYSE-) HYSEQ INC.
XX PI Drmanac RT, Liu C, Tang YT;
XX DR WPI; 2001-639362/73.
XX DR P-PSDB; ABG20022.
XX PT New isolated polynucleotide and encoded polypeptides, useful in
XX PT diagnostics, forensics, gene mapping, identification of mutations
XX PT responsible for genetic disorders or other traits and to assess
XX PT biodiversity.
XX PS Claim 1; SEQ ID No 20013; 103pp; English.
XX CC The invention relates to isolated polynucleotide (I) and
XX CC polypeptide (II) sequences. (I) is useful as hybridisation probes,
XX CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
XX CC and gene mapping, and in recombinant production of (II). The
XX CC polynucleotides are also used in diagnostics as expressed sequence tags
XX CC for identifying expressed genes. (I) is useful in gene therapy techniques
XX CC to restore normal activity of (II) or to treat disease states involving
XX CC (II). (II) is useful for generating antibodies against it, detecting or
XX CC quantitating a polypeptide in tissue, as molecular weight markers and as
XX CC a food supplement. (II) and its binding partners are useful in medical
XX CC imaging of sites expressing (II). (I) and (II) are useful for treating
XX CC disorders involving aberrant protein expression or biological activity.
XX CC The polypeptide and polynucleotide sequences have applications in
XX CC diagnostics, forensics, gene mapping, identification of mutations
XX CC responsible for genetic disorders or other traits to assess biodiversity
XX CC and to produce other types of data and products dependent on DNA and
XX CC amino acid sequences. AAS64197-AAS94564 represent novel human
XX CC diagnostic coding sequences of the invention.
XX CC Note: The sequence data for this patent did not appear in the printed
XX CC specification, but was obtained in electronic format directly from WIPO
XX CC at ftp.wipo.int/pub/published_pcl_sequences.
XX SQ Sequence 8294 BP; 2476 A; 1994 C; 1774 G; 2050 T; 0 other;
XX
XX Query Match 100.0%; Score 99; DB 23; Length 8294;
XX Best Local Similarity 100.0%; Pred. No. 8.3e-26;
XX Matches 99; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 TCACCCCTGAGCGCAAGCAATGTTGGCATCTGTAAAGACCACTAGATCCAGCA 60
XX ||||||||||||||||||||||||||||||||||||||||||||||||||||||
XX DB 4652 TCACCCCTGAGCGCAAGCAATGTTGGCATCTGTAAAGACCACTAGATCCAGCA 4711
XX
XX QY 61 GCCCAGACCCCTTTCTTTGTGTCAAGAAAGCGGGGAAA 99
XX ||||||||||||||||||||||||||||||||||||||||||||||||||||||
XX DB 4712 GCCCAGACCCCTTTCTTTGTGTGTCAAGAAAGCGGGGAAA 4750
XX
XX RESULT 12
XX ABN97929
XX ID ABN97929 standard; DNA; 10499 BP.
XX AC ABN97929;
XX XX
XX DT 01-AUG-2002 (first entry)
XX XX
XX DE Human retroviral sequence HERV-7q.
XX XX
```

```
KW Autoimmune disease; HERV-7q; chromosome 7q; immunotherapy;
KM multiple sclerosis; ds.
XX OS
XX OS Human retrovirus.
XX PN WO967395-A1.
XX PD 29-DEC-1999.
XX PF 23-JUN-1999; 99WO-FR01513.
XX PR 23-JUN-1998; 98FR-0007920.
XX PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.
XX PI Alliel PM, Perin J, Rieger F;
XX DR WPI; 2000-160587/14.
XX PT New nucleic acid sequences of human endogenous retrovirus, HERV-7q,
XX PT used for diagnosis, treatment and prevention of autoimmune and
XX PT neurological diseases.
XX PS Claim 3; Fig 1; 225pp; French.
XX CC The present invention relates to new nucleic acid sequences of human
XX CC endogenous retrovirus, HERV-7q, which is located on chromosome 7q.
XX CC Regulatory elements associated with HERV-7q may alter expression of other
XX CC genes (even remote genes) on the same chromosome, inducing immunological
XX CC and/or neurological changes (which may be pathological or protective/
XX CC curative). HERV-7q peptides can be used to improve efficiency of the
XX CC immune response, e.g. in immunotherapy. HERV-7q peptides and their coding
XX CC sequences can be used in immunogenic or vaccinating compositions, for
XX CC protection against autoimmune diseases, particularly multiple sclerosis.
XX CC The peptides may also be used (by sequence comparison) to detect/identify
XX CC endogenous retroviruses that are abnormally expressed in cancer,
XX CC neuropathologies or other autoimmune diseases. The present sequence was
XX CC used to illustrate the invention.
XX SQ Sequence 10499 BP; 3048 A; 2676 C; 2280 G; 2495 T; 0 other;
XX
XX Query Match 100.0%; Score 99; DB 21; Length 10499;
XX Best Local Similarity 100.0%; Pred. No. 9e-26;
XX Matches 99; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 1 TCACCCCTGAGCGCAAGCAATGTTGGCATCTGTAAAGACCACTAGATCCAGCA 60
XX ||||||||||||||||||||||||||||||||||||||||||||||||||||||
XX DB 2502 TCACCCCTGAGCGCAAGCAATGTTGGCATCTGTAAAGACCACTAGATCCAGCA 2561
XX
XX QY 61 GCCCAGACCCCTTTCTTTGTGTCAAGAAAGCGGGGAAA 99
XX ||||||||||||||||||||||||||||||||||||||||||||||||||||||
XX DB 2562 GCCCAGACCCCTTTCTTTGTGTGTCAAGAAAGCGGGGAAA 2600
XX
XX RESULT 13
XX ABL61744
XX ID ABL61744 standard; DNA; 56093 BP.
XX AC ABL61744;
XX XX
XX DT 15-MAY-2002 (first entry)
XX XX
XX DE Colon adenocarcinoma related gene sequence SEQ ID NO:81.
XX KW Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;
XX KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;
XX KW cytostatic; gene therapy; antineoplastic; Wilms' tumour; adenocarcinoma;
XX KW gene; ds.
XX OS
XX OS Homo sapiens.
XX PN WO200194629-A2.
XX XX
```

PD 13-DEC-2001.  
XX  
PF 30-MAY-2001; 2001MO-US10838.  
XX  
PR 05-JUN-2000; 2000US-209473P.  
PR 05-JUN-2000; 2000US-209531P.  
PR 18-SEP-2000; 2000US-233133P.  
PR 18-SEP-2000; 2000US-233617P.  
PR 20-SEP-2000; 2000US-234009P.  
PR 20-SEP-2000; 2000US-234034P.  
PR 20-SEP-2000; 2000US-234052P.  
PR 22-SEP-2000; 2000US-234509P.  
PR 22-SEP-2000; 2000US-234567P.  
PR 25-SEP-2000; 2000US-234923P.  
PR 25-SEP-2000; 2000US-234924P.  
PR 25-SEP-2000; 2000US-235077P.  
PR 25-SEP-2000; 2000US-235082P.  
PR 25-SEP-2000; 2000US-235134P.  
PR 25-SEP-2000; 2000US-235280P.  
PR 26-SEP-2000; 2000US-235637P.  
PR 26-SEP-2000; 2000US-235638P.  
PR 27-SEP-2000; 2000US-235711P.  
PR 27-SEP-2000; 2000US-235720P.  
PR 27-SEP-2000; 2000US-235840P.  
PR 27-SEP-2000; 2000US-235863P.  
PR 28-SEP-2000; 2000US-236028P.  
PR 28-SEP-2000; 2000US-236032P.  
PR 28-SEP-2000; 2000US-236033P.  
PR 28-SEP-2000; 2000US-236034P.  
PR 28-SEP-2000; 2000US-236109P.  
PR 28-SEP-2000; 2000US-236111P.  
PR 29-SEP-2000; 2000US-236842P.  
PR 29-SEP-2000; 2000US-236891P.  
PR 02-OCT-2000; 2000US-237172P.  
PR 02-OCT-2000; 2000US-237173P.  
PR 02-OCT-2000; 2000US-237278P.  
PR 02-OCT-2000; 2000US-237294P.  
PR 02-OCT-2000; 2000US-237295P.  
PR 02-OCT-2000; 2000US-237316P.  
PR 03-OCT-2000; 2000US-237425P.  
PR 03-OCT-2000; 2000US-237598P.  
PR 03-OCT-2000; 2000US-237604P.  
PR 03-OCT-2000; 2000US-237606P.  
PR 03-OCT-2000; 2000US-237608P.  
PR 01-NOV-2000; 2000US-244867P.  
PR 01-NOV-2000; 2000US-245084P.  
XX  
PA (AVAL-) AVALON PHARM.  
XX  
PI Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;  
PI Soppet DR, Weaver Z;  
XX  
XX WPI; 2002-188264/24.  
XX  
PT Screening for anti-neoplastic agent involves exposing cells to a  
PT chemical agent to be tested for anti-neoplastic activity, and  
PT determining a change in expression of a gene of a signature gene set -  
XX  
PS Claim 1; SEQ ID 81; 44pp; English.  
XX  
XX The present invention describes a method (M1) for screening for an  
XX anti-neoplastic agent. The method involves exposing cells to a chemical  
XX agent to be tested for anti-neoplastic activity, determining a change in  
XX expression of at least one gene (I) of a signature gene set, where (I)  
XX comprises a sequence (S) selected from 8447 sequences (given in ABL61664  
XX to ABL70110), or is at least 95% identical to (S), where a change in  
XX expression is indicative of anti-neoplastic activity. (I) has cyclostatic  
XX activity and can be used in gene therapy. M1 can be used for screening  
XX an anti-neoplastic agent, and can be used for producing a product which  
XX is the data collected with respect to the anti-neoplastic agent as a  
XX result of M1, and the data is sufficient to convey the chemical  
XX structure and/or properties of the agent. M1 can be used in the  
XX treatment of cancer such as colon, breast, stomach, lung, thyroid.

CC oesophageal, ovarian, kidney, prostate or pancreatic cancer,  
CC adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,  
CC infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine  
CC carcinoma, papillary carcinoma and Wilms' tumour.  
XX  
SQ Sequence 56093 BP; 16164 A; 12346 C; 10702 G; 16681 T; 0 other;  
Query Match 100.0%; Score 99; DB 24; Length 56093;  
Best Local Similarity 100.0%; Pred. No. 1,6e-25;  
Matches 99; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 TCACCCCTGAGCGCAAGGCATGTGGCATGCTGTAAAGCACTAGATCCAGCA 60  
DB 30502 TCACCCCTGAGCGCAAGGCATGTGGCATGCTGTAAAGCACTAGATCCAGCA 30561  
QY 61 GCCCAGACCCCTTCTTGTCAGAGAAAGCGGGA 99  
DB 30562 GCCCAGACCCCTTCTTGTCAGAGAAAGCGGGA 30600  
RESULT 14  
AA572234  
ID AA572234 standard; cDNA; 1687 BP.  
XX  
AC AA572234;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE DNA encoding novel human diagnostic protein #8038.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX  
OS Homo sapiens.  
XX  
PN W0200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001MO-US08631.  
XX  
PR 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;  
PI  
XX  
XX WPI; 2001-639362/73.  
XX  
XX P-PSDB; ABC08047.  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
XX  
XX Claim 1; SEQ ID No 8038; 103pp; English.  
XX  
XX The invention relates to isolated polynucleotide (I) and  
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,  
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
XX and gene mapping, and in recombinant production of (II). The  
XX polynucleotides are also used in diagnostics as expressed sequence tags  
XX for identifying expressed genes. (I) is useful in gene therapy techniques  
XX to restore normal activity of (II) or to treat disease states involving  
XX (II). (II) is useful for generating antibodies against it, detecting or  
XX quantitating a polypeptide in tissue, as molecular weight markers and as  
XX a food supplement. (II) and its binding partners are useful in medical  
XX imaging of sites expressing (II). (I) and (II) are useful for treating  
XX disorders involving aberrant protein expression or biological activity.  
XX The polypeptide and polynucleotide sequences have applications in  
XX diagnostics, forensics, gene mapping, identification of mutations  
XX responsible for genetic disorders or other traits to assess biodiversity



CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 1687 BP; 513 A; 434 C; 357 G; 383 T; 0 other;  
Query Match 98.4%; Score 97.4; DB 23; Length 1687;  
Best Local Similarity 99.0%; Pred. No. 1.8e-25;  
Matches 98; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
OY 1 TCACCCCTGAGCGCAAGCAATGTTGGCATGCTGTAAGAGACCCTAGATCCAGCA 60  
DB 143 TCACCCCTGAGCGCAAGCAATGTTGGCATGCTGTAAGAGACCCTAGATCCAGCA 202  
OY 61 GCCCAGACCCCTTCTTGTGTGTCAGAAAGGCGGGGAAA 99  
DB 203 GCCCAGACCCCTTCTTGTGTGTCAGAAAGGCGGGGAAA 241  
RESULT 15  
AAS77018  
ID AAS77018 standard; cDNA; 1729 BP.  
XX  
AC AAS77018;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE DNA encoding novel human diagnostic protein #12822.  
XX  
KW Human; Chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US08631.  
XX  
PR 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;  
XX  
DR WPI: 2001-639362/73.  
DR P-PSDB; ABG12831.  
XX  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity  
PS Claim 1; SEQ ID No 12822; 103pp; English.  
XX  
XX The invention relates to isolated polynucleotide (I) and  
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,  
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
XX and gene mapping, and in recombinant production of (II). The  
XX polynucleotides are also used in diagnostics as expressed sequence tags  
XX for identifying expressed genes. (I) is useful in gene therapy techniques  
XX to restore normal activity of (II) or to treat disease states involving  
XX (II). (II) is useful for generating antibodies against it, detecting or  
XX quantitating a polypeptide in tissue, as molecular weight markers and as  
XX a food supplement. (II) and its binding partners are useful in medical  
XX imaging of sites expressing (II). (I) and (II) are useful for treating  
XX disorders involving aberrant protein expression or biological activity.  
XX The polypeptide and polynucleotide sequences have applications in

CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 1729 BP; 525 A; 429 C; 362 G; 413 T; 0 other;  
Query Match 98.4%; Score 97.4; DB 23; Length 1729;  
Best Local Similarity 99.0%; Pred. No. 1.8e-25;  
Matches 98; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
OY 1 TCACCCCTGAGCGCAAGCAATGTTGGCATGCTGTAAGAGACCCTAGATCCAGCA 60  
DB 143 TCACCCCTGAGCGCAAGCAATGTTGGCATGCTGTAAGAGACCCTAGATCCAGCA 202  
OY 61 GCCCAGACCCCTTCTTGTGTGTCAGAAAGGCGGGGAAA 99  
DB 203 GCCCAGACCCCTTCTTGTGTGTCAGAAAGGCGGGGAAA 241  
Search completed: May 2, 2003, 11:29:37  
Job time : 177.717 secs



GenCore version 5.1.5  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 2, 2003, 11:21:20 ; Search time 1140.21 Seconds

(without alignments)  
1406.195 Million cell updates/sec

Title: US-09-719-554-3\_COPY\_2502\_2600

Perfect score: 99

Sequence: 1 tcaccctgagcgcaagagc.....tgtrcaagaagcgagga 99

Scoring table: IDENTITY\_NUC

Gapop 10.0, Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :  
EST:  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vit:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_tod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the total score being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	93.2	94.1	432	17	AQ438690
2	93.2	94.1	481	17	AQ412419
3	92.6	93.5	373	12	BF883154
4	89.4	90.3	620	17	AQ624662
5	86.2	87.1	565	9	AL698433
6	86.2	87.1	592	17	AQ465425

7	83.6	84.4	561	9	AA307753
8	82.2	83.0	538	17	AQ122695
9	81.4	82.2	441	17	B84492
10	81.4	82.2	632	17	B84492
11	53.6	54.1	624	17	AQ014046
12	47.2	47.7	515	17	AQ060950
13	34.2	34.5	386	14	T96504
14	31.6	31.9	482	14	B0560733
15	31.6	31.9	683	14	BM941798
16	31.6	31.9	722	13	B1851147
17	31.3	31.3	816	17	BH565798
18	30.8	31.1	918	14	BQ070614
19	30.4	30.7	659	17	BH107401
20	30.3	30.3	664	9	BT112104
21	29.6	29.9	188	9	A1877491
22	29.6	29.9	310	14	BQ974577
23	29.6	29.9	511	14	BQ262033
24	29.4	29.7	572	17	BH738095
25	29.4	29.7	688	17	BH731212
26	29.4	29.7	707	17	BH532111
27	29.4	29.7	756	17	BH703188
28	29.2	29.5	380	12	BG572101
29	28.6	28.9	742	13	B1888821
30	28.6	28.9	765	13	BG972789
31	28.4	28.7	720	12	BE872347
32	28.5	28.5	544	17	A2957805
33	28.3	28.3	506	17	AQ148423
34	28.3	28.3	623	17	BH516961
35	28.3	28.3	658	13	B1891305
36	27.8	28.1	223	17	BH599181
37	27.8	28.1	399	10	AW156589
38	27.8	28.1	418	17	AQ045384
39	27.8	28.1	487	17	AQ126708
40	27.8	28.1	535	17	AQ599844
41	27.8	28.1	546	17	BH672523
42	27.8	28.1	572	17	AQ504740
43	27.8	28.1	575	17	BH478321
44	27.8	28.1	577	17	BH248929
45	27.8	28.1	581	17	AQ389315

## ALIGNMENTS

RESULT 1  
AQ438690/c  
LOCUS  
DEFINITION  
HS.5141.A2.F08.SPEE.RPCT-11 Human Male BAC Library Homo sapiens  
ACCESSION  
AQ438690  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
human.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
1 (bases 1 to 432)  
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,  
Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and  
Hood,L.  
Sequence-tagged connectors: A sequence approach to mapping and  
scanning the human genome  
Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)  
99380589  
Contact: Mahairas GG, Wallace JC, Hood L  
High Throughput Sequencing Center  
University of Washington  
401 Queen Anne Avenue North, Seattle, WA 98109, USA  
Tel: (206) 616-3618  
Fax: (206) 616-3887  
Email: jwallace@u.washington.edu  
Clones are derived from the human BAC library RPCT-11. For BAC  
library availability, please contact Pieter de Jong

(pieter@dejong.med.buffalo.edu). Clones may be purchased from BACpac Resources (<http://bacpac.med.buffalo.edu/ordering/bac.htm>) or from Research Genetics (<http://www.htsc.washington.edu>). BAC end Web Server: <http://www.htsc.washington.edu>  
Plate: 717 row: K column: 16  
Seq primer: SP6  
Class: BAC ends  
High quality sequence stop: 432.

## FEATURES

Location/Qualifiers

1. 432

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_1lb="RPCI-11 Human Male BAC Library"

/sex="Male"

/note="Vector: pBAC3.6; Site.1: EcoRI; Site.2: EcoRI; Male blood DNA was isolated from one randomly chosen donor and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBAC3.6 vector at EcoRI sites"

## BASE COUNT

107 a 104 c 89 g 131 t 1 others

## Query Match

Best Local Similarity 96.9%; Pred. No. 1.2e-20; Length 432;

Matches: 95; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TCACCCCTGAGCGCAAGCAATGTTGGCATGCTGTAAAGACCACTAGATCCAGCA 60

DB 326 TCACCCCTGAGCGCAAGCAATGTTGGCATGCTGTAAAGACCACTAGATCCAGCA 267

QY 61 GCCCAGACCCCTTTCTTTGTGTCAGAAAGCGCGGAA 98

DB 266 GCCCAGACCCCTTTCTTTGTGTCAGAAAGCGCGGAA 229

## RESULT 2

LOCUS A0412419

DEFINITION RPCI-11-170F15.TU RPCI-11 Homo sapiens genomic clone RPCI-11-170F15

ACCESSION A0412419

VERSION A0412419.1 GI:4471907

KEYWORDS GSS:

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

Zhuo, S., Adams, M.D., Niernan, W., Malek, J., de Jong, P. and Venter, J.C.

TITLE

Use of BAC End Sequences from Library RPCI-11 for Sequence-Ready Map Building

JOURNAL

COMMENT

Unpublished (1997)  
Other GSSs: RPCI-11-170F15.TV  
Contact: Shaying Zhao, William Niernan, Mark Adams  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: hbeet@ig.org  
Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACpac Resources (<http://bacpac.med.buffalo.edu/ordering>) or from Research Genet cs (<http://www.htsc.washington.edu>). BAC end search page: [http://www.tigr.org/tldb/humgen/bac\\_end\\_search/bac\\_end\\_search.html](http://www.tigr.org/tldb/humgen/bac_end_search/bac_end_search.html).  
Seq primer: SP6  
Class: BAC ends.

## FEATURES

Location/Qualifiers

1. 481

/organism="Homo sapiens"

/db\_xref="GDB:7565030"  
/db\_xref="taxon:9606"  
/clone\_1lb="RPCI-11-170F15"  
/clone\_1lb="RPCI-11"  
/sex="Male"  
/cell\_type="lymphocytes"  
/note="Vector: pBAC3.6; Site.1: EcoRI; Site.2: EcoRI; RPCI11 Human Male BAC Library"

## BASE COUNT

122 a 111 c 106 g 140 t 2 others

## Query Match

Best Local Similarity 94.1%; Score 93.2; DB 17; Length 481;

Matches: 95; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 TCACCCCTGAGCGCAAGCAATGTTGGCATGCTGTAAAGACCACTAGATCCAGCA 60

DB 340 TCACCCCTGAGCGCAAGCAATGTTGGCATGCTGTAAAGACCACTAGATCCAGCA 281

QY 61 GCCCAGACCCCTTTCTTTGTGTCAGAAAGCGCGGAA 98

DB 280 GCCCAGACCCCTTTCTTTGTGTCAGAAAGCGCGGAA 243

## RESULT 3

LOCUS BF883154

DEFINITION QV3-ET0211-071200-529-c02 ET0211 Homo sapiens CDNA, mRNA sequence.

ACCESSION BF883154

VERSION BF883154.1 GI:12273280

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Brites, M.R., Nagel, M.A., da Silva, W., Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldmann, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H., Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.

Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

JOURNAL

COMMENT

Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICP Human Cancer Genome Project. This entry can be seen in the following URL  
(<http://www.ludwig.org.br/scripts/gethtml2.pl?tl-QV3-ET0211-071200-529-c02&f3-2000-12-07&f4-1>)  
Seq primer: puc 18 forward  
High quality sequence start: 34  
High quality sequence stop: 373.

## FEATURES

source

Location/Qualifiers

1. 373

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_1lb="ET0211"

/dev stage="Adult"

/note="Organ: lung tumor; Vector: puc18; Site.1: SmaI; Site.2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under

BASE COUNT 108 a 94 c 96 g 74 t 1 others  
 ORIGIN

Query Match . 93.5%; Score 92.6; DB 12; Length 373;  
 Best Local Similarity 96.0%; Pred. No. 1.8e-20;  
 Matches .95; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 TCACCCCTGAGCGCAAGGCAATGTTGGCATGCTGTAAGACCACTAGATCCAGCA 60  
 DB 78 TCACCCCTGAGCGCAAGGCAATGTTGGCATGCTGTAAGACCACTAGATCCAGCA 137  
 QY 61 GCCCAGACCCCTTTCTTTGTGCTCAAGAAAGCGGGAAA 99  
 DB 138 GCCCGATCCCTTTCTTTGTGCTCAAGAAAGCGGGAAA 176

## RESULT 4

LOCUS A0624662/c 620 bp DNA linear GSS 16-JUN-1999  
 DEFINITION CITBI-EI-2657A12.TR CITBI-EI Homo sapiens genomic clone 2657A12,  
 DNA sequence.

ACCESSION A0624662 GI:5087054  
 VERSION A0624662  
 KEYWORDS GSS.

## SOURCE

ORGANISM

Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 620)  
 Zhao,S., Adams,M.D., Nierman,M., Malek,J., Shizuya,H., Simon,M. and  
 Venter,J.C.

## AUTHORS

## TITLE

Use of BAC End Sequences from Caltech Libraries for Sequence-Ready  
 Map Building

## JOURNAL

Unpublished (1997)

Other GSSs: CITBI-EI-2657A12.TF

Contact: Shaying Zhao, William Nierman, Mark Adams

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850

Tel: 301 838 0200

Fax: 301 838 0208

Email: hbe@tigr.org

Clones are available from Research Genetics (info@resgen.com). BAC  
 end search page:

http://www.tigr.org/tdb/humgen/bac\_end\_search/bac\_end\_search.html.

Seq primer: M13 Reverse

Class: BAC ends.

## FEATURES

Location/Qualifiers

1..620

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="2657A12"

/clone\_1ib="CITBI-EI"

/sex="male"

/cell\_type="sperm"

/note="Vector: pBeloBAC11; Site\_1: EcoRI; Site\_2: EcoRI;  
 Caltech Human BAC Library D"

BASE COUNT 170 a 135 c 120 g 195 t

## ORIGIN

Query Match 90.3%; Score 89.4; DB 17; Length 620;  
 Best Local Similarity 93.9%; Pred. No. 2.4e-19;  
 Matches 93; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 TCACCCCTGAGCGCAAGGCAATGTTGGCATGCTGTAAGACCACTAGATCCAGCA 60  
 DB 320 TCACCCCTGAGCGCAAGGCAATGTTGGCATGCTGTAAGACCACTAGATCCAGCA 261

QY 61 GCCCAGACCCCTTTCTTTGTGCTCAAGAAAGCGGGAAA 99  
 DB 260 GCCCATACCCCTTTCTTTGTGCTCAAGAAAGCGGGAAA 222

RESULT 5  
 LOCUS AL698433/c 565 bp mRNA linear EST 21-MAR-2002

DEFINITION DKFZp686K02108.r1 686 (synonym: hicc3) Homo sapiens cDNA clone  
 DKFZp686K02108.5, mRNA sequence.

ACCESSION AL698433  
 VERSION AL698433.1 GI:19618973  
 KEYWORDS EST.

## SOURCE

ORGANISM

human.

Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 565)  
 Bloeker,H., Boecker,M., Brandt,P., Mewes,W., Well,B. and Wiemann  
 S.

## AUTHORS

## TITLE

EST (Bloeker,H., Boecker,M., Brandt,P., Mewes,W., Well,B. and  
 Wiemann,S.)

Unpublished (1999)

Contact: Bloeker H

## COMMENT

MIPS

Am Klopferspitz 18a D-82152 Martinsried, Germany

This is the 5' sequence of the clone insert

Clone from S. Wiemann, Molecular Genome Analysis, German Cancer  
 Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;

sequenced by GBF (National Research Centre for Biotechnology Ltd.,  
 Braunschweig/Germany) within the cDNA sequencing consortium of the  
 German Genome Project.

No sl sequence available.

This clone (DKFZp686K02108) is available at the RZPD in Berlin.  
 Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059  
 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.

## FEATURES

Location/Qualifiers

1..565

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="DKFZp686K02108"

/clone\_1ib="686 (synonym: hicc3)"

/tissue\_type="human skeletal muscle"

/dev\_stage="adult"

/lab\_host="DH10B"

/note="Vector: pTR1p1ex2; Site\_1: SfiI; Site\_2: SfiIb;  
 cDNA-collection"

BASE COUNT 142 a 126 c 117 g 180 t

## ORIGIN

Query Match 87.1%; Score 86.2; DB 9; Length 565;  
 Best Local Similarity 91.9%; Pred. No. 2.7e-18;  
 Matches 91; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 1 TCACCCCTGAGCGCGCAAGGCAATGTTGGCATGCTGTAAGACCACTAGATCCAGCA 60  
 DB 516 TCACCCCTGAGCGCGCAAGGCAATGTTGGCATGCTGTAAGACCACTAGATCCAGCA 457

QY 61 GCCCAGACCCCTTTCTTTGTGCTCAAGAAAGCGGGAAA 99  
 DB 456 GCCCAGACCCCTTTCTTTGTGCTCAAGAAAGCGGGAAA 418

## RESULT 6

LOCUS A0465425/c 592 bp DNA linear GSS 23-APR-1999  
 DEFINITION HS\_5107\_B2.C02.SP6E RPCT-11 Human Male BAC Library Homo sapiens  
 genomic clone Plate=683 Col=4 Row=F, DNA sequence.

ACCESSION A0465425  
 VERSION A0465425.1 GI:4642520  
 KEYWORDS GSS.

## SOURCE

ORGANISM

Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 592)  
 Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,

COMMENT: Contact: Mahairas GG, Wallace JC, Hood L  
High Throughput Sequencing Center

University of Washington  
401 Queen Anne Avenue North, Seattle, WA 98109, USA  
Tel: (206) 616-3618  
Fax: (206) 616-3887  
Email: jwallace@u.washington.edu  
Sequence Tagged Connector  
Plate: 3080 row: G column: 8  
Class: BAC ends

High quality sequence stop: 538.

## FEATURES

source

Location/Qualifiers  
1. .538

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="plate:3080 Col=8 Row=G"  
/clone\_lib="CIT Approved Human Genomic Sperm Library D"  
/sex="male"  
/note="Organ: sperm; Vector: pBelobAC11; BAC clones in E-Coli DH10B"  
BASE COUNT 125 a 131 c 117 g 163 t 2 others  
ORIGIN

Query Match 83.0%; Score 82.2; DB 17; Length 538;  
Best Local Similarity 91.6%; Pred. No. 5.8e-17;  
Matches 87; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

OY 5 CCGTAGGCGCAAGGCAATGTTGGCATGCTGTAAGGACCACTAGATCAGACGCC 64

DB 374 CTTGAGCTCAAGGCAATGTTGGCATGCTGTAAGGACCACTAGATCAGACGCC 315

OY 65 AGACCCCTTCTTGTGCTCAAGAAAGCGGGA 99

DB 314 GGACCCCTTCTTGTGCTCAAGAAAGCGGGA 280

## RESULT 9

B84492/c

LOCUS

DEFINITION B84492 441 bp DNA linear GSS 09-APR-1999  
RPc111-2404.TP RPc11-11 Homo sapiens genomic clone RPc11-2404, DNA

ACCESSION

VERSION B84492  
KEYWORDS GSS.  
SOURCE GI:2925624

ORGANISM

human.  
Homo sapiens

REFERENCE

AUTHORS

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 441)  
Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K.,  
Goider,K., Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and  
Venter,J.C.

Use of BAC End Sequences for Sequence-Ready Map Building (1998)  
Unpublished (1998)  
Contact: Mark Adams  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: mdadams@tigr.org

TITLE

JOURNAL

COMMENT

Clones are derived from the human BAC library RPc11-11. For BAC  
library availability, please contact Pieter de Jong  
(pieter@dejong.med.buffalo.edu). Clones may be purchased from  
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from  
Research Genetics (info@resgen.com). BAC end search page:  
http://www.tigr.org/tldb/humgen/bac\_end\_search/bac\_end\_search.html  
Seq primer: SP6  
Class: BAC ends.

## FEATURES

source

Location/Qualifiers  
1. .441

/organism="Homo sapiens"  
/db\_xref="GDB:7509171"  
/db\_xref="taxon:9606"  
/clone="RPc11-2404"

/clone\_lib="RPc11-11"  
/sex="Male"  
/cell\_type="Lymphocytes"  
/note="Vector: pBAC3.6; Site\_1: EcoRI; Site\_2: EcoRI;  
RPc111 Human Male BAC library"  
BASE COUNT 106 a 108 c 94 g 133 t  
ORIGIN

Query Match 82.2%; Score 81.4; DB 17; Length 441;  
Best Local Similarity 88.9%; Pred. No. 1e-16;  
Matches 88; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

OY 1 TCACCCGAGCGCAAGGCAATGTTGGCATGCTGTAAGGACCACTAGATCAGACA 60

DB 307 TCATCTCTGACTGAAAGGCAATGTTGAGAAATGCTGGTAAGGACCACTAGATCAGACA 248

OY 61 GCCCAGACCCCTTCTTGTGCTCAAGAAAGCGGGA 99

DB 247 ACCCAGACCCCTTCTTGTGCTCAAGAAAGGGA 209

## RESULT 10

A0014046/c

LOCUS

DEFINITION A0014046 632 bp DNA linear GSS 14-APR-1999  
RPc111-2404.TKBR RPc11-11 Homo sapiens genomic clone RPc11-2404,  
DNA sequence.

ACCESSION

VERSION A0014046  
KEYWORDS GSS.  
SOURCE GI:3186611

ORGANISM

human.  
Homo sapiens

REFERENCE

AUTHORS

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 632)  
Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K.,  
Goider,K., Berry,K., Granger,D., Suh,E., Wible,C., de Jong,P. and  
Venter,J.C.

Use of BAC End Sequences for Sequence-Ready Map Building (1998)  
Unpublished (1998)  
Other-GSSs: RPc111-2404.TP

JOURNAL

COMMENT

Contact: Mark Adams  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: mdadams@tigr.org  
Clones are derived from the human BAC library RPc11-11. For BAC  
library availability, please contact Pieter de Jong  
(pieter@dejong.med.buffalo.edu). Clones may be purchased from  
BACPAC Resources (http://bacpac.med.buffalo.edu/ordering) or from  
Research Genetics (info@resgen.com). BAC end search page:  
http://www.tigr.org/tldb/humgen/bac\_end\_search/bac\_end\_search.html  
Class: BAC ends.

## FEATURES

source

Location/Qualifiers  
1. .632

/organism="Homo sapiens"  
/db\_xref="GDB:7509171"  
/db\_xref="taxon:9606"  
/clone="RPc11-2404"  
/clone\_lib="RPc11-11"  
/sex="Male"  
/cell\_type="Lymphocytes"  
/note="Vector: pBAC3.6; Site\_1: EcoRI; Site\_2: EcoRI;  
RPc111 Human Male BAC library"  
BASE COUNT 156 a 144 c 133 g 199 t  
ORIGIN

Query Match 82.2%; Score 81.4; DB 17; Length 632;  
Best Local Similarity 88.9%; Pred. No. 1.1e-16;  
Matches 88; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

OY 1 TCACCCGAGCGCAAGGCAATGTTGGCATGCTGTAAGGACCACTAGATCAGACA 60

DB 321 TCACCTCTGACCTGAAGCAATGTGACATGCTGTAGGACACAGATCCAGCA 262  
OY 61 GCCACAGACCCCTTCTTGTGTGTCAGAAAGCGGAAA 99  
DB 261 ACCCAGACCCCTTCTTGTGTGTCAGAAAGAGGAAA 223

RESULT 11  
A0060950 624 bp DNA linear GSS 31-JUL-1998  
LOCUS A0060950  
DEFINITION CIT-HSP-2347H19.TF CIT-HSP Homo sapiens genomic clone 2347H19, DNA sequence.  
ACCESSION A0060950  
VERSION A0060950.1 GI:3362862  
KEYWORDS GSS.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 624)  
Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K., Golden,K., Berry,K., Granger,D., Sun,E., Wible,C., Shizuya,H., Simon,M. and Venter,J.C.  
Use of a random BAC End Sequence Database for Sequence-Ready Map Building (1998)  
Unpublished (1998)  
Other GSSs: CIT-HSP-2347H19.TR  
Contact: Mark Adams  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: mdadams@tigr.org  
Clones are available from Research Genetics (info@resgen.com). BAC end search page:  
[http://www.tigr.org/tdb/hungen/bac\\_end\\_search/bac\\_end\\_search.html](http://www.tigr.org/tdb/hungen/bac_end_search/bac_end_search.html)  
Seq primer: M13-21  
Class: BAC ends

FEATURES  
Location/Qualifiers  
source 1..624  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="2347H19"  
/clone\_1lb="CIT-HSP"  
/sex="Male"  
/cell\_type="Sperm"  
/note="Vector: pBelBAC11, site\_1: HindIII, site\_2: HindIII"

BASE COUNT 244 a 130 c 125 g 125 t  
ORIGIN

Query Match 54.1%; Score 53.6; DB 17; Length 624;  
Best Local Similarity 93.3%; Pred. No. 2.2e-07;  
Matches 56; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

OY 40 AACGACACTAGATCCAGACGACCCCTTCTTGTGTGTCAGAAAGCGGAAA 99  
DB 448 AAAAATACCTAGATCCAGACGACCCCTTCTTGTGTGTCAGAAAGCGGAAA 507

RESULT 12  
A0609568 515 bp DNA linear GSS 15-JUN-1999  
LOCUS A0609568  
DEFINITION HS\_5081.B1.D08.T7A RPCI-11 Human Male BAC Library Homo sapiens genomic clone Plate-657 Col-15 Row-H, DNA sequence.  
ACCESSION A0609568  
VERSION A0609568.1 GI:5070844  
KEYWORDS GSS.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE 1 (bases 1 to 515)  
Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L.  
Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome  
Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)  
99380589  
CONTACT: Mahairas GC, Wallace JC, Hood L  
High Throughput Sequencing Center  
University of Washington  
401 Queen Anne Avenue North, Seattle, WA 98109, USA  
Tel: (206) 616-3887  
Fax: (206) 616-3887  
Email: jwallace@u.washington.edu  
Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieter@dejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources ([http://bacpac.med.buffalo.edu/ordering\\_bac.htm](http://bacpac.med.buffalo.edu/ordering_bac.htm)) or from Research Genetics (info@resgen.com). BAC end Web Server: <http://www.htsc.washington.edu>  
Plate: 657 row: H column: 15  
Seq primer: T7  
Class: BAC ends  
High quality sequence stop: 515.

FEATURES  
Location/Qualifiers  
source 1..515  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="Plate-657 Col-15 Row-H"  
/clone\_1lb="RPCI-11 Human Male BAC library"  
/sex="Male"  
/note="Vector: pBAC3.6; Site\_1: EcoRI; Site\_2: EcoRI; Male blood DNA was isolated from one randomly chosen donor and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBAC3.6 vector at EcoRI sites"

BASE COUNT 140 a 114 c 118 g 117 t 26 others  
ORIGIN

Query Match 47.7%; Score 47.2; DB 17; Length 515;  
Best Local Similarity 72.6%; Pred. No. 2.8e-05;  
Matches 61; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

OY 13 GCAAGGCAATGTGGCATGCTGTAAAGACCACTAGATCCAGACGACCCCT 72  
DB 138 GCGAAGCTATAGACATGCTGTAAAGACCACTAGATCCAGACGACCCCT 197

OY 73 TTCTTGTGTGTCAGAAAGCGGG 96  
DB 198 TTTTTCGCGCTAAAGAGTGA 221

RESULT 13  
T96504 386 bp mRNA linear EST 27-MAR-1995  
LOCUS T96504  
DEFINITION yec34g02.r1 Stratagene lung (#937210) Homo sapiens cDNA clone IMAGE:119666 5', mRNA sequence.  
ACCESSION T96504  
VERSION T96504.1 GI:735128  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 386)  
Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chapell,B., Chisoe,S., Dietrich,N., Dubuque,T., Favello,A., Gish,W., Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N., Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierly-Meg,J., Trevaaks,E., Underwood,K., Wohlmann,P., Waterston,R., Wilson,R. and Marra,M.



TITLE Generation and analysis of 280,000 human expressed sequence tags  
JOURNAL Genome Res. 6 (9), 807-828 (1996)  
MEDLINE 97044478  
COMMENT Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: estewatson.wustl.edu

Insert Size: 1072  
High quality sequence stops: 254 Source: IMAGE Consortium, LNL  
This clone is available royalty-free through LNL; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
Insert Length: 1072 Std Error: 0.00  
Seq primer: M13RPI  
High quality sequence stop: 254.

FEATURES  
source

1. 386  
Location/Qualifiers  
/organism="Homo sapiens"  
/db\_xref="GDB:487955"  
/db\_xref="taxon:9606"  
/clone="IMAGE:119666"  
/clone\_id="Stratagene lung (#937210)"  
/sex="male"  
/dev\_stage="72 years"  
/lab\_host="SOLR cells (kanamycin resistant)"  
/note="Organ: lung; Vector: phuescript SK-; Site\_1: EcorI  
; Site\_2: XhoI; Cloned unidirectionally. Primer: Oligo  
dT, normal lung. Average insert size: 1.0 kb; Uni-ZAP XR  
Vector; -5' adaptor sequence: 5' GAATTCGACACGAG 3' -3'  
adaptor sequence: 5' CTCGAGTCTTTTCTTTTCTTTT 3' "

BASE COUNT 108 a 83 c 102 g 90 t 3 others  
ORIGIN

Query Match 34.5%; Score 34.2; DB 14; Length 386;  
Best Local Similarity 92.3%; Pred. No. 0.59;  
Matches 36; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 61 GCCCAGACCCTTCTTGTGTCAAGAAAGCGCGGAA 99  
|||||  
Db 1 GCCCAGACCCTTCTTGTGTCAAGAGGTGGGAA 39

RESULT 14  
BO560733/c 482 bp mRNA linear EST 20-JUN-2002

DEFINITION BO560733 H4066B08-5 NIA Mouse 7.4k cDNA clone Set Mus musculus cDNA clone  
H4066B08 5', mRNA sequence.

ACCESSION BO560733  
VERSION BO560733.1 GI:21461618  
KEYWORDS EST  
SOURCE house mouse.  
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 482)  
AUTHORS VanBuren V., Piao Y., Dudekula D.B., Qian Y., Carter M.G., Martin  
P.R., Stagg C.A., Bassey U., Alpa K., Hamatani T., Karpi G.J.,  
Luo A.G. and Ko M.S.H.

TITLE Assembly, verification, and initial annotation of NIA 7.4k mouse  
cDNA clone set

JOURNAL Unpublished (2002)  
COMMENT Contact: Yong Qian

Laboratory of Genetics  
National Institute on Aging/National Institutes of Health  
333 Cassell Drive, Suite 3000, Baltimore, MD 21224-6820, USA  
Email: cdna@igsun.grc.nia.nih.gov

This clone set has been freely distributed to the community. Please  
visit [http://igsun.grc.nia.nih.gov/cDNA/NIA\\_7\\_4k.html](http://igsun.grc.nia.nih.gov/cDNA/NIA_7_4k.html) for details.

Seq primer: -21M13 Reverse  
High quality sequence stop: 482

POLYA-No.

FEATURES  
source

Location/Qualifiers  
1. 482  
/organism="Mus musculus"  
/strain="C57BL/6"  
/db\_xref="taeST:H4066B08-5"  
/db\_xref="taxon:10090"  
/clone="H4066B08"  
/clone\_id="NIA Mouse 7.4k cDNA clone Set"  
/sex="mixed"  
/dev\_stage="mixed"  
/lab\_host="DH10B"  
/note="Vector: pSPORT1; Site\_1: SalI; Site\_2: NotI; This  
clone is among a rearrayed set of 7,407 clones from more  
than 20 cDNA libraries."

BASE COUNT 117 a 124 c 126 g 115 t  
ORIGIN

Query Match 31.9%; Score 31.6; DB 14; Length 482;  
Best Local Similarity 62.8%; Pred. No. 4.6;  
Matches 49; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

OY 12 CCAGAGCAATGTTGGCATGCTGTAAAGACCACTAGATCCAGACCCAGACCC 71  
|||||

Db 165 CCAGAGCAATGTTGGCATGCTGTAAAGACCACTAGATCCAGACCCAGACCC 106  
|||||

OY 72 TTTCTTGTGTCAAGAA 89  
|||||  
Db 105 TTTCTTGTGTCAAGAA 88

RESULT 15  
BM941798/c

DEFINITION BM941798 683 bp mRNA linear EST 29-APR-2002  
UT-M-CG0p-bq1-c-04-0-UI.r1 NIH\_BMAP\_Ret4\_S2 Mus musculus cDNA clone  
UT-M-CG0p-bq1-c-04-0-UI 5', mRNA sequence.

ACCESSION BM941798  
VERSION BM941798.1 GI:19400950  
KEYWORDS EST  
SOURCE house mouse.  
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniala; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
1 (bases 1 to 683)  
AUTHORS Bonaldo M.F., Lennon G. and Soares M.B.

TITLE Normalization and subtraction: two approaches to facilitate gene  
discovery

JOURNAL Genome Res. 6 (9), 791-806 (1996)  
MEDLINE 97044477  
COMMENT Contact: Chin, H  
National Institute of Mental Health  
6001 Executive Blvd. Room 7N-7190, MSC 9643, Bethesda, MD  
20892-9643, USA  
Tel: 301 443 1706  
Fax: 301 443 9890  
Email: mestr@mail.nih.gov

Tissue Procurement: Dr. Xin-Yuan Fu, Yale University School of  
Medicine

cDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Researchers may obtain clones from Research  
Genetics (www.resgen.com).

Seq primer: M13 REVERSE.  
Location/Qualifiers

1. 683  
Location/Qualifiers

/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"

/clone="UT-M-CG0p-bq1-c-04-0-UI"  
/clone\_id="NIH\_BMAP\_Ret4\_S2"

/lab\_host="DH10B (Life Technologies)"  
/note="Vector: pRT30-Pac (Pharmacia) with a modified

polylinker. Site\_1: Not I; Site\_2: Eco RI; The



GenCore version 5.1.5  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 2, 2003, 11:30:30 ; Search time 39.6 Seconds  
(without alignments)  
766.692 Million cell updates/sec

Title: US-09-719-554-3\_COPY\_2502\_2600

Perfect score: 99  
Sequence: 1 tcacccctgagcgcaaacg.....tgctcaagaagcgaggaa 99

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_NA:\*  
1: /cgn2\_6/ptodata/2/ina/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/2/ina/6A.COMB.seq:\*  
3: /cgn2\_6/ptodata/2/ina/6A.COMB.seq:\*  
4: /cgn2\_6/ptodata/2/ina/6B.COMB.seq:\*  
5: /cgn2\_6/ptodata/2/ina/PCrus.COMB.seq:\*  
6: /cgn2\_6/ptodata/2/ina/Backflist.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	31.8	32.1	1100	4	US-08-456-200B-17
2	31.8	32.1	2470	1	US-07-745-206A-14
3	31.8	32.1	2470	2	US-08-311-363-14
4	31.8	32.1	5467	1	US-07-745-206A-12
5	31.8	32.1	5467	2	US-08-311-363-12
6	31.8	32.1	7175	1	US-08-455-543A-8
7	31.8	32.1	7175	2	US-08-193-078B-8
8	31.8	32.1	7175	2	US-08-223-305C-8
9	31.8	32.1	7175	2	US-08-149-097D-8
10	31.8	32.1	7175	3	US-08-948-386-8
11	31.8	32.1	7175	3	US-08-450-562-8
12	31.8	32.1	7175	4	US-08-984-709A-8
13	31.8	32.1	7175	4	US-08-450-272-8
14	31.8	32.1	7177	4	US-09-268-163-7
15	31.8	32.1	7266	3	US-08-713-118-1
16	31.8	32.1	7266	3	US-09-452-007-1
17	31.8	32.1	7362	1	US-08-455-543A-7
18	31.8	32.1	7362	2	US-08-193-078B-7
19	31.8	32.1	7362	2	US-08-223-305C-7
20	31.8	32.1	7362	3	US-08-149-097D-7
21	31.8	32.1	7362	3	US-08-948-386-7
22	31.8	32.1	7362	4	US-08-450-562-7
23	31.8	32.1	7362	4	US-08-984-709A-7
24	31.8	32.1	7362	4	US-08-450-272-7
25	31.8	32.1	7364	4	US-09-268-163-5
26	31.8	32.1	7376	4	US-09-268-163-3
27	30.3	30.3	7011	4	US-09-268-163-9

28	26.8	27.1	4403765	4	US-09-103-840A-2	Sequence 2, Appl
29	26.8	27.1	4411529	4	US-09-103-840A-1	Sequence 1, Appl
30	26.6	26.9	590	1	US-08-580-038-8	Sequence 8, Appl
31	26.4	26.7	6038	4	US-09-305-639-4	Sequence 4, Appl
32	26.4	26.7	7622	4	US-09-305-639-1	Sequence 1, Appl
33	26.3	26.3	3117	3	US-08-909-954-3	Sequence 3, Appl
34	25.8	26.1	4136	4	US-09-103-875-2	Sequence 2, Appl
35	25.6	25.9	162450	4	US-09-345-882-1	Sequence 1, Appl
36	25.4	25.7	2301	4	US-09-448-218D-1	Sequence 1, Appl
37	25.4	25.7	2301	4	US-09-448-218D-3	Sequence 3, Appl
38	25.4	25.7	2301	4	US-09-449-218D-5	Sequence 5, Appl
39	25.4	25.7	2301	4	US-09-449-218D-7	Sequence 7, Appl
40	25.4	25.7	9301	4	US-09-449-218D-18	Sequence 18, Appl
41	25.2	25.5	138	3	US-08-812-121-17	Sequence 17, Appl
42	25.2	25.5	138	4	US-09-403-672-17	Sequence 17, Appl
43	25	25.3	595	1	US-08-580-038-25	Sequence 25, Appl
44	25	25.3	2004	1	US-08-471-033-18	Sequence 18, Appl
45	25	25.3	2004	2	US-08-471-044-18	Sequence 18, Appl

## ALIGNMENTS

RESULT 1  
US-08-456-200B-17  
Sequence 17, Application US/08456200B  
Patent No. 6229000  
GENERAL INFORMATION:  
APPLICANT: Franz, Jurgen; Weingartner, Bernhard;  
APPLICANT: Unterbeck, Axel; Rae, Peter  
TITLE OF INVENTION: TISSUE-SPECIFIC HUMAN NEURONL  
TITLE OF INVENTION: CALCIUM CHANNEL SUB-TYPES AND  
THEIR USE  
NUMBER OF SEQUENCES: 17  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: SPRUNG HORN KRAMER & WOODS  
STREET: 660 White Plains Road  
CITY: Tarrytown  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10591-5144  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 MB  
MEDIUM TYPE: Storage  
COMPUTER: NEC Powermate SX/20  
OPERATING SYSTEM: DOS  
SOFTWARE: WordPerfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/456,200B  
FILING DATE: 31-MAY-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/094,712  
FILING DATE: 19-JUL-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/858,278  
FILING DATE: 26-MAR-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/064,778  
FILING DATE: 19-MAY-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: DE 41 10 785  
FILING DATE: 04-APR-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Kurt G. Briscoe  
REGISTRATION NUMBER: 33,141  
REFERENCE/DOCKET NUMBER: Bayer 8398.3-KGB  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (914) 332-1700  
TELEFAX: (914) 332-1844  
INFORMATION FOR SEQ ID NO: 17:  
SEQUENCE CHARACTERISTICS:

LENGTH: 1100 nucleotides  
TYPE: Nucleotide  
STRANDEDNESS: Single  
TOPOLOGY: Linear  
MOLECULE TYPE: cDNA  
US-08-456-200B-17

Query Match 32.1%; Score 31.8; DB 4; Length 1100;  
Best Local Similarity 61.4%; Pred. No. 0.015;  
Matches 51; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 9 GAGCGCAAGGCAATGTTGGCATGCTGTAAGAGACCACTAGATCCAGACCCAGAC 68  
||| ||||| | ||||| ||| ||| ||| |||  
DB 340 GATCTTCAAGCGGAGCAATGCTGCGCGAGAGAGAGAGAGATGACAGAGAGAGTC 399

OY 69 CCCTTCTTGTGTCAGAAAG 91  
||||| ||| ||||| ||  
DB 400 CCCTTGGACGTGCTGAAGAGAG 422

## RESULT 2

US-07-745-206A-14  
Sequence 14, Application US/07745206A  
Patent No. 5428921  
GENERAL INFORMATION:  
APPLICANT: Harpold, Michael  
APPLICANT: Ellis, Steven  
APPLICANT: Williams, Mark  
APPLICANT: McCue, Ann  
APPLICANT: Feldman, Daniel  
TITLE OF INVENTION: Human Calcium Channel Compositions and  
METHODS  
NUMBER OF SEQUENCES: 32  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fitch, Even, Tabin & Flannery  
STREET: 135 S. LaSalle  
CITY: Chicago  
STATE: Illinois  
COUNTRY: U.S.A.  
ZIP: 60603  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07745,206A  
FILING DATE: 19910815  
CLASSIFICATION: 435.  
ATTORNEY/AGENT INFORMATION:  
NAME: Feder, Scott B  
REFERENCE/DOCKET NUMBER: 51504  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-372-7842  
INFORMATION FOR SEQ. ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2470 base pairs  
TYPE: NUCLEIC ACID  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..2469  
US-07-745-206A-14

Query Match 32.1%; Score 31.8; DB 1; Length 2470;  
Best Local Similarity 61.4%; Pred. No. 0.021;  
Matches 51; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 9 GAGCGCAAGGCAATGTTGGCATGCTGTAAGAGACCACTAGATCCAGACCCAGAC 68  
||| ||||| | ||||| ||| ||| ||| |||  
DB 1173 GATCTTCAAGCGGAGCAATGCTGCGCGAGAGAGAGAGATGACAGAGAGAGTC 1232

OY 69 CCCTTCTTGTGTCAGAAAG 91  
||||| ||| ||||| ||  
DB 1233 CCCTTGGACGTGCTGAAGAGAG 1255

## RESULT 3

US-08-311-363-14  
Sequence 14, Application US/08311363  
Patent No. 5876958  
GENERAL INFORMATION:  
APPLICANT: Harpold, Michael  
APPLICANT: Ellis, Steven  
APPLICANT: Williams, Mark  
APPLICANT: Feldman, Daniel  
APPLICANT: McCue, Ann  
APPLICANT: Brenner, Robert  
TITLE OF INVENTION: Human Calcium Channel Compositions and  
METHODS  
NUMBER OF SEQUENCES: 32  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Brown, Martin, Haller & McClain  
STREET: 1660 Union Street  
CITY: San Diego  
STATE: California  
COUNTRY: USA  
ZIP: 92101-2926  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/311,363  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/745,206  
FILING DATE: 15-AUG-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Seidman, Stephanie L.  
REGISTRATION NUMBER: 33,779  
REFERENCE/DOCKET NUMBER: 6362-51506  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619)238-0999  
TELEFAX: (619)238-0062  
INFORMATION FOR SEQ. ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2470 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..2469  
US-08-311-363-14

Query Match 32.1%; Score 31.8; DB 2; Length 2470;  
Best Local Similarity 61.4%; Pred. No. 0.021;  
Matches 51; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 9 GAGCGCAAGGCAATGTTGGCATGCTGTAAGAGACCACTAGATCCAGACCCAGAC 68  
||| ||||| | ||||| ||| ||| ||| |||  
DB 1173 GATCTTCAAGCGGAGCAATGCTGCGCGAGAGAGAGAGATGACAGAGAGAGTC 1232  
OY 69 CCCTTCTTGTGTCAGAAAG 91  
||||| ||| ||||| ||  
DB 1233 CCCTTGGACGTGCTGAAGAGAG 1255

## RESULT 4

US-07-745-206A-12  
Sequence 12, Application US/07745206A

Patent No. 5429921  
GENERAL INFORMATION:  
APPLICANT: Harpold, Michael  
APPLICANT: Ellis, Steven  
APPLICANT: Williams, Mark  
APPLICANT: McCue, Ann  
TITLE OF INVENTION: Human Calcium Channel Compositions and  
METHODS  
NUMBER OF SEQUENCES: 32  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fitch, Even, Tabin & Flannery  
STREET: 135 S. LaSalle  
CITY: Chicago  
STATE: Illinois  
COUNTRY: U.S.A.  
ZIP: 60603  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/745,206A  
FILING DATE: 19910815  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Feder, Scott B  
REFERENCE/DOCKET NUMBER: 51504  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 312-372-7842  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5467 base pairs  
TYPE: NUCLEIC ACID  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: join(144..3164, 3168..3245, 3249..3386, 3390  
LOCATION: ..3392, 3396..3488, 3495..3539, 3543..3581, 3585  
LOCATION: ..3587, 3591..3626, 3630..3693, 3693..3737, 3744  
LOCATION: ..3746, 3750..4823, 4827..4841, 4845..5006, 5010  
LOCATION: ..5096, 5100..5306, 5310..5366, 5370..5465)  
US-07-745-206A-12  
Query Match 32.1%; Score 31.8; DB 1; Length 5467;  
Best Local Similarity 61.4%; Pred. No. 0.03;  
Matches 51; Conservative 0; Mismatches 32; Indels 0; Gaps 0;  
OY 9 GAGCCCAAGGCAATGTTGGCATCTGTAAGACCACTAGATCCAGACCCAGAC 68  
DB 1316 GATCTTCAAGCGGAGGAAGTATGTCGCCGAGGAGACAGAGATGCAAGAGATC 1375  
OY 69 CCCTTCTTTGTGTCAGAAAG 91  
DB 1376 CCCTTGGACGTCTGAAGAGAG 1398  
RESULT 5  
US-08-311-363-12  
Sequence 12, Application US/08311363  
Patent No. 587658  
GENERAL INFORMATION:  
APPLICANT: Harpold, Michael  
APPLICANT: Ellis, Steven  
APPLICANT: Williams, Mark  
APPLICANT: Feldman, Daniel  
APPLICANT: McCue, Ann  
APPLICANT: Brenner, Robert  
TITLE OF INVENTION: Human Calcium Channel Compositions and  
METHODS

NUMBER OF SEQUENCES: 32  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Brown, Martin, Haller & McClain  
STREET: 1660 Union Street  
CITY: San Diego  
STATE: California  
COUNTRY: USA  
ZIP: 92101-2926  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/311,363  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/745,206  
FILING DATE: 15-AUG-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Seidman, Stephanie L.  
REGISTRATION NUMBER: 33,779  
REFERENCE/DOCKET NUMBER: 6362-51506  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619)238-0999  
TELEFAX: (619)238-0062  
INFORMATION FOR SEQ ID NO: 12:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5467 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: join(144..3164, 3168..3245, 3249..3386, 3390  
LOCATION: ..3392, 3396..3488, 3495..3539, 3543..3581, 3585  
LOCATION: ..3587, 3591..3626, 3630..3689, 3693..3737, 3744  
LOCATION: ..3746, 3750..4823, 4827..4841, 4845..5006, 5010  
LOCATION: ..5096, 5100..5306, 5310..5366, 5370..5465)  
US-08-311-363-12  
Query Match 32.1%; Score 31.8; DB 2; Length 5467;  
Best Local Similarity 61.4%; Pred. No. 0.03;  
Matches 51; Conservative 0; Mismatches 32; Indels 0; Gaps 0;  
OY 9 GAGCCCAAGGCAATGTTGGCATCTGTAAGACCACTAGATCCAGACCCAGAC 68  
DB 1316 GATCTTCAAGCGGAGGAAGTATGTCGCCGAGGAGACAGAGATGCAAGAGATC 1375  
OY 69 CCCTTCTTTGTGTCAGAAAG 91  
DB 1376 CCCTTGGACGTCTGAAGAGAG 1398  
RESULT 6  
US-08-455-543A-8  
Sequence 8, Application US/08455543A  
Patent No. 5792846  
GENERAL INFORMATION:  
APPLICANT: Harpold, Michael  
APPLICANT: Ellis, Steven  
APPLICANT: Williams, Mark  
APPLICANT: Feldman, Daniel  
APPLICANT: McCue, Ann  
APPLICANT: Brenner, Robert  
TITLE OF INVENTION: HUMAN CALCIUM CHANNEL COMPOSITIONS AND  
METHODS  
NUMBER OF SEQUENCES: 57  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Brown, Martin, Haller & McClain  
STREET: 1660 Union Street  
CITY: San Diego

```

STATE: California
COUNTRY: USA
ZIP: 92101-2926
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/455.543A
FILING DATE: May 31, 1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/223.305
FILING DATE: April 4, 1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/868.354
FILING DATE: April 10, 1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/745.206
FILING DATE: 15-AUG-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/620.250
FILING DATE: 30-NOV-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/482.384
FILING DATE: 20-FEB-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/603.751
FILING DATE: 04-APR-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO PCT/US89/01408
FILING DATE: 04-APR-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/176.899
FILING DATE: 04-APR-1988
ATTORNEY/AGENT INFORMATION:
NAME: Seligman, Stephanie L.
REGISTRATION NUMBER: 33,779
REFERENCE/DOCKET NUMBER: 6362-52517
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619)238-0999
TELEFAX: (619)238-0062
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 7175 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 144..6857
FEATURE:
NAME/KEY: 5'UTR
LOCATION: 1..143
FEATURE:
NAME/KEY: 3'UTR
LOCATION: 6855..7175
US-08-455-543A-8
Query Match 32.1%; Score 31.8; DB 1; Length 7175;
Best Local Similarity 61.4%; Pred. No. 0.033; Mismatches 32; Indels 0; Gaps 0.
Matches 51; Conservative 0;
QY 9 GAGCGCAAGGCATGTTGGCGATCTGTGTAAGGACCACTGAAATCCAGCAGCCAGAC 68
Db 1316 GATCTTCAAGCGGAGGAAGATCATCTCGTGCCGAGAGGACAGCAATGCAGAGGAAGTC 1375
QY 69 CCCCTTCTTTGTGTGTCAGGAAG 91
Db 1376 CCCCTTGGACGTGCTGAAGAGAG 1398

```

```

      1  RESULT 7
      2  US-08-193-078B-8
      3  ; Sequence 8, Application US/08193078B
      4  ; Patent No. 5846757
      5  ; GENERAL INFORMATION:
      6  ; APPLICANT: Harpold, Michael
      7  ; APPLICANT: Ellis, Steven
      8  ; APPLICANT: Williams, Mark
      9  ; APPLICANT: Feldman, Daniel
     10  ; APPLICANT: McCue, Ann
     11  ; APPLICANT: Brenner, Robert
     12  ; TITLE OF INVENTION: HUMAN CALCIUM CHANNEL COMPOSITIONS AND
     13  ; TITLE OF INVENTION: METHODS
     14  ; NUMBER OF SEQUENCES: 29
     15  ; CORRESPONDENCE ADDRESS:
     16  ; ADDRESSEE: BROWN, MARTIN, HALLER & MCCLAIN
     17  ; STREET: 1660 UNION STREET
     18  ; CITY: SAN DIEGO
     19  ; STATE: CA
     20  ; COUNTRY: USA
     21  ; ZIP: 92101
     22  ;
     23  ; COMPUTER READABLE FORM:
     24  ; MEDIUM TYPE: Floppy disk
     25  ; COMPUTER: IBM PC compatible
     26  ; OPERATING SYSTEM: PC-DOS/MS-DOS
     27  ; SOFTWARE: Patent Release #1.0, Version #1.25
     28  ; CURRENT APPLICATION DATA:
     29  ; APPLICATION NUMBER: US/08/193.078B
     30  ; FILING DATE: 07-FEB-1994
     31  ; CLASSIFICATION: 435
     32  ; PRIOR APPLICATION DATA:
     33  ; APPLICATION NUMBER: US 07/868,354
     34  ; FILING DATE: 10-APR-1992
     35  ; PRIOR APPLICATION DATA:
     36  ; APPLICATION NUMBER: US 07/745,206
     37  ; FILING DATE: 15-AUG-1991
     38  ; ATTORNEY/AGENT INFORMATION:
     39  ; NAME: Seidman, Stephanie L.
     40  ; REGISTRATION NUMBER: 33,779
     41  ; REFERENCE/DOCKET NUMBER: 6362-53607
     42  ; TELECOMMUNICATION INFORMATION:
     43  ; TELEPHONE: 619-238-0062
     44  ; TELEFAX: 619-238-0062
     45  ; INFORMATION FOR SEQ ID NO: 8:
     46  ; SEQUENCE CHARACTERISTICS:
     47  ; LENGTH: 7175 base pairs
     48  ; TYPE: nucleic acid
     49  ; STRANDEDNESS: double
     50  ; TOPOLOGY: linear
     51  ; MOLECULE TYPE: DNA (genomic)
     52  ; FEATURE:
     53  ; NAME/KEY: CDS
     54  ; LOCATION: 144..6857
     55  ; FEATURE:
     56  ; NAME/KEY: 5'UTR
     57  ; LOCATION: 1..143
     58  ; FEATURE:
     59  ; NAME/KEY: 3'UTR
     60  ; LOCATION: 6855..7175
     61  ;
     62  ; US-08-193-078B-8
     63  ;
     64  ; Query Match
     65  ; Best Local Similarity 32.1%; Score 31.8; DB 2; Length 7175;
     66  ; Matches 51; Conservativity 61.4%; Pred. No.0.033;
     67  ; Mismatches 37; Indels 0; Gaps 0;
     68  ;
     69  ; 9 GAGCGCAAGGCAATGTTGGCGCATCGTGTAAGGACCACTAGAAATCCAGCAGCCAGAC 68
     70  ; 11 11111 11 11111111 1111 11 111111 11 11
     71  ; Db 1316 GATCTTCAAGGGGAGGAATGATCTGCTGAGCAGAGAGACAGAGATGCAAGAGAGAATGTC 1375
     72  ; 111111 111 11111 11
     73  ;
     74  ; 69 CCCTTCTTTTGTGGCTCAGAAGAAG 91
     75  ; 111111 111 11111 11
     76  ;
     77  ; Db 1376 CCCTTGGACGTCGTGAAGAGAG 1398

```

```

1      RESULT 8
2      US-08-223-305C-8
3      Sequence 8, Application US/08223305C
4      Patent No. 5851824
5      GENERAL INFORMATION:
6      APPLICANT: Harpold, Michael
7      APPLICANT: Ellis, Steven
8      APPLICANT: Williams, Mark
9      APPLICANT: Feldman, Daniel
10     APPLICANT: McCue, Ann
11     APPLICANT: Brenner, Robert
12     TITLE OF INVENTION: HUMAN CALCIUM CHANNEL COMPOSITIONS AND
13     TITLE OF INVENTION: METHODS
14     NUMBER OF SEQUENCES: 57
15     CORRESPONDENCE ADDRESS:
16     ADDRESSEE: Brown, Martin, Haller & McClain
17     STREET: 1660 Union Street
18     CITY: San Diego
19     STATE: California
20     COUNTRY: USA
21     ZIP: 92101-2926
22     COMPUTER READABLE FORM:
23     MEDIUM TYPE: Diskette
24     COMPUTER: IBM Compatible
25     OPERATING SYSTEM: DOS
26     SOFTWARE: FastSeq Version 1.5
27     CURRENT APPLICATION DATA:
28     APPLICATION NUMBER: US/08/223,305C
29     FILING DATE: April 4, 1994
30     PRIOR APPLICATION DATA:
31     APPLICATION NUMBER: 07/868,354
32     FILING DATE: April 10, 1992
33     PRIOR APPLICATION DATA:
34     APPLICATION NUMBER: US 07/745,206
35     FILING DATE: 15-AUG-1991
36     PRIOR APPLICATION DATA:
37     APPLICATION NUMBER: 07/620,250
38     FILING DATE: 30-NOV-1990
39     PRIOR APPLICATION DATA:
40     APPLICATION NUMBER: US 07/482,384
41     FILING DATE: 20-FEB-1990
42     PRIOR APPLICATION DATA:
43     APPLICATION NUMBER: 07/603,751
44     FILING DATE: 04-APR-1989
45     PRIOR APPLICATION DATA:
46     APPLICATION NUMBER: US 07/176,899
47     FILING DATE: 04-APR-1989
48     PRIOR APPLICATION DATA:
49     APPLICATION NUMBER: WO PCT/US89/01408
50     FILING DATE: 04-APR-1989
51     PRIOR APPLICATION DATA:
52     APPLICATION NUMBER: US 07/176,899
53     FILING DATE: 04-APR-1989
54     ATTORNEY/AGENT INFORMATION:
55     NAME: Seidman, Stephanie L.
56     REGISTRATION NUMBER: 33,779
57     REFERENCE/DOCKET NUMBER: 55216 (P519739)
58     TELECOMMUNICATION INFORMATION:
59     TELEPHONE: (619)238-0999
60     TELEFAX: (619)238-0062
61     INFORMATION FOR SEQ ID NO: 8:
62     SEQUENCE CHARACTERISTICS:
63     LENGTH: 7175 base pairs
64     TYPE: nucleic acid
65     STRANDEDNESS: double
66     TOPOLOGY: linear
67     MOLECULE TYPE: DNA (genomic)
68     FEATURE:
69     NAME/KEY: CDS
70     LOCATION: 144..6857
71     FEATURE:
72     NAME/KEY: 5'UTR
73     LOCATION: 1..143
74     FEATURE:
75     NAME/KEY: 3'UTR

```

```

! LOCATION: 6855..7175
US-08-223-305C-8

Query Match      32.1%; Score 31.8; DB 2; Length 7175;
Best Local Similarity 61.4%; Pred. No. 0.033;
Matches 51; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY    9  GAGCGCAAAAGGCAATGGTTGGGCATCTGGTAAAGGACCACTGAATCCAGACCAGAC 68
      ||| |||| | | ||||||| ||| | |||| ||| | |||
Db    1316 GATCTTC AAGCGGGAGGAGCATCATCTGGCCGAGGAGCACGAATGCCAGAGAGACA GTC 1375

QY    69  CCCCTTTCTTTGGTGTCAGAAGAAG 91
      ||||| |||| | ||| ||| ||| |||
Db    1376 CCGTTTGACGCGTCTGAAGAAGAG 1398
```

RESULT 9  
US-08-149-097D-8  
Sequence 8, Application US/08149097D  
Patent No. 5874236  
GENERAL INFORMATION:  
APPLICANT: Harpold, Michael  
APPLICANT: Ellis, Steven  
APPLICANT: Williams, Mark  
APPLICANT: Feldman, Daniel  
APPLICANT: McQue, Ann  
TITLE OF INVENTION: HUMAN CALCIUM CHANNEL COMPOSITIONS AND  
TITLE OF INVENTION: METHODS  
NUMBER OF SEQUENCES: 40  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Brown, Martin, Haller & McClain  
STREET: 1660 Union Street  
CITY: San Diego  
STATE: California  
COUNTRY: USA  
ZIP: 92101-2926  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/149, 097D  
FILING DATE: 05-NOV-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/105, 536  
FILING DATE: 11-AUG-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/US92/06903  
FILING DATE: 14-AUG-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/914, 231  
FILING DATE: 13-JUL-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/868, 354  
FILING DATE: 10-APR-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/745, 206  
FILING DATE: 15-AUG-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/620, 250  
FILING DATE: 30-NOV-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/482, 384  
FILING DATE: 20-FEB-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/603, 751  
FILING DATE: 04-APR-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/US89/01408  
FILING DATE: 04-APR-1989

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/176,899  
FILING DATE: 04-APR-1988  
ATTORNEY/AGENT INFORMATION:  
NAME: Seidman, Stephanie L.  
REGISTRATION NUMBER: 33,779  
REFERENCE/DOCKET NUMBER: 6362-55038  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 238-0999  
TELEFAX: (619) 238-0062  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7175 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 144..6857  
FEATURE:  
NAME/KEY: 5'UTR  
LOCATION: 1..143  
FEATURE:  
NAME/KEY: 3'UTR  
LOCATION: 6855..7175  
US-08-149-097D-8

Query Match 32.1%; Score 31.8; DB 2; Length 7175;  
Best Local Similarity 61.4%; Pred. No. 0.033;  
Matches 51; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 9 GACCGAAGCAGATGTTGGCATGCTGTAAGGACCACTAGATCCAGACCCGAGAC 68  
DB 1316 GATCTTCAGAGCGGAGAGATCATGCTGCCGAGAGAGACAGAAATGCAGAGAAAGTC 1375  
QY 69 CCCTTCTTGTGTCAGAAAG 91  
DB 1376 CCCTTGGACGTCTGTAAGAGAG 1398

RESULT 10  
US-08-949-386-8  
Sequence 8, Application US/08949386  
Patent No. 6090623  
GENERAL INFORMATION:  
APPLICANT: Harpold, Michael  
APPLICANT: Ellis, Steven  
APPLICANT: Williams, Mark  
APPLICANT: McCue, Ann  
APPLICANT: Gillespie, Alison  
TITLE OF INVENTION: HUMAN CALCIUM CHANNEL COMPOSITIONS AND  
METHODS  
NUMBER OF SEQUENCES: 38  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Brown, Martin, Haller & McClain  
STREET: 1660 Union Street  
CITY: San Diego  
STATE: California  
COUNTRY: US  
ZIP: 92101  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/08/290,012  
FILING DATE: 11-AUG-1994

APPLICATION NUMBER: 08/149,097  
FILING DATE: 5-NOV-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/105,536  
FILING DATE: 11-AUG-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Seidman, Stephanie L.  
REGISTRATION NUMBER: 33,779  
REFERENCE/DOCKET NUMBER: 519808  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 238-0999  
TELEFAX: (619) 238-0062  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7175 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 144..6857  
FEATURE:  
NAME/KEY: 5'UTR  
LOCATION: 1..143  
FEATURE:  
NAME/KEY: 3'UTR  
LOCATION: 6855..7175  
US-08-949-386-8

Query Match 32.1%; Score 31.8; DB 3; Length 7175;  
Best Local Similarity 61.4%; Pred. No. 0.033;  
Matches 51; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 9 GACCGAAGCAGATGTTGGCATGCTGTAAGGACCACTAGATCCAGACCCGAGAC 68  
DB 1316 GATCTTCAGAGCGGAGAGATCATGCTGCCGAGAGAGACAGAAATGCAGAGAAAGTC 1375  
QY 69 CCCTTCTTGTGTCAGAAAG 91  
DB 1376 CCCTTGGACGTCTGTAAGAGAG 1398

RESULT 11  
US-08-450-562-8  
Sequence 8, Application US/08450562  
Patent No. 6096514  
GENERAL INFORMATION:  
APPLICANT: Harpold, Michael  
APPLICANT: Ellis, Steven  
APPLICANT: Williams, Mark  
APPLICANT: McCue, Ann  
APPLICANT: Gillespie, Alison  
APPLICANT: Feldman, Daniel  
APPLICANT: Brenner, Robert  
TITLE OF INVENTION: HUMAN CALCIUM CHANNEL COMPOSITIONS AND  
METHODS  
NUMBER OF SEQUENCES: 38  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Brown, Martin, Haller & McClain  
STREET: 1660 Union Street  
CITY: San Diego  
STATE: California  
COUNTRY: US  
ZIP: 92101  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
FILING DATE:  
APPLICATION NUMBER: US/08/450,562



1	CLASSIFICATION: 435
2	PRIOR APPLICATION DATA:
3	APPLICATION NUMBER: 08/404,950
4	FILING DATE: 13-MAR-1995
5	APPLICATION NUMBER: 08/336,257
6	FILING DATE: 7-NOV-1994
7	PRIOR APPLICATION DATA:
8	APPLICATION NUMBER: 08/314,083
9	FILING DATE: 28-SEPT-1994
10	PRIOR APPLICATION DATA:
11	APPLICATION NUMBER: 08/311,363
12	FILING DATE: 23-SEPT-1994
13	PRIOR APPLICATION DATA:
14	APPLICATION NUMBER: 08/290,012
15	FILING DATE: 11-AUG-1994
16	PRIOR APPLICATION DATA:
17	APPLICATION NUMBER: 08/223,305
18	FILING DATE: 4-APR-1994
19	PRIOR APPLICATION DATA:
20	APPLICATION NUMBER: 08/193,078
21	FILING DATE: 07-FEB-1994
22	PRIOR APPLICATION DATA:
23	APPLICATION NUMBER: 08/149,097
24	FILING DATE: 5-NOV-1993
25	PRIOR APPLICATION DATA:
26	APPLICATION NUMBER: 08/105,536
27	FILING DATE: 11-AUG-1993
28	PRIOR APPLICATION DATA:
29	APPLICATION NUMBER: 07/914,231
30	FILING DATE: 13-JULY-1992
31	PRIOR APPLICATION DATA:
32	APPLICATION NUMBER: 07/868,354
33	FILING DATE: 10-APR-1992
34	PRIOR APPLICATION DATA:
35	APPLICATION NUMBER: PCT/US92/06903
36	FILING DATE: 14-AUG-1992
37	PRIOR APPLICATION DATA:
38	APPLICATION NUMBER: 07/745,206
39	FILING DATE: 15-AUG-1991
40	PRIOR APPLICATION DATA:
41	APPLICATION NUMBER: 07/620,250
42	FILING DATE: 30-NOV-1990
43	PRIOR APPLICATION DATA:
44	APPLICATION NUMBER: 07/603,751
45	FILING DATE: 08-NOV-1990
46	PRIOR APPLICATION DATA:
47	APPLICATION NUMBER: 07/482,384
48	FILING DATE: 02-FEB-1990
49	PRIOR APPLICATION DATA:
50	APPLICATION NUMBER: PCT/US89/01408
51	FILING DATE: 04-APR-1989
52	PRIOR APPLICATION DATA:
53	APPLICATION NUMBER: 07/176,899
54	FILING DATE: 04-APR-1988
55	ATTORNEY/AGENT INFORMATION:
56	NAME: Seidman, Stephanie L.
57	REGISTRATION NUMBER: 33,779
58	REFERENCE/DOCKET NUMBER: 6362-519812
59	TELECOMMUNICATION INFORMATION:
60	TELEPHONE: (619) 238-0999
61	TELEFAX: (619) 238-0062
62	INFORMATION FOR SEQ ID NO: 8:
63	SEQUENCE CHARACTERISTICS:
64	LENGTH: 7115 base pairs
65	TYPE: nucleic acid
66	STRANDEDNESS: double
67	TOPOLOGY: linear
68	MOLECULE TYPE: DNA (genomic)
69	FEATURE:
70	NAME/KEY: CDS
71	LOCATION: 144..6857
72	FEATURE:
73	NAME/KEY: 5'UTR

```

; LOCATION: 1..143
;
; FEATURE:
; NAME/KEY: 3'UTR
; LOCATION: 6855..7175
;
US-08-450-562-8

```

Query Match	32.1%;	Score 31.8;	DB 3;	length 7175;
Best Local Similarity	61.4%;	Pred. No. 0.033;		
Matches 51;	Conservative	0;	Mismatches 32;	Indels 0;
				Gaps 0;

Qy 9 GACCGCAAGGCAATGTTGGGCATCTGGTAAAGGACCACTGAAATCCAGACCCACAG 68  
||| ||||| | ||||| ||| | ||||| ||| | |||  
Db 1316 GATCTTCAAGCGCGAGGAAGTCAATCTGGCCGAGGAGACAGCAAGTCAAGAGAGAAAGTTC 1375

Qy 69 CCCTTCTTGTGGTCAAGAAAG 91  
 ||||| ||| | ||||| ||  
 Db 1376 CCCTTGGACGTGCTGAAGAGAG 1398

RESULT 12  
US-08-984-709A-8

; sequence 8, Application US/08984/09A  
; Patent No. 6320032  
; Patent No. 6320032

```
; GENERAL INFORMATION:
; APPLICANT: Williams, Mark E.
```

APPLICANT: Stauderman, Kenneth A.  
APPLICANT: Harpold, Michael M.

TITLE OF INVENTION: HUMAN CALCIUM CHANNEL COMPOSITIONS AND METHODS

```
; NUMBER OF SEQUENCES: 52  
; CORRESPONDENCE ADDRESS:  
ADDRESS
```

ADDRESSEE: Heller Erman White & McCaulliffe  
STREET: 4250 Executive Square, Suite 700  
CITY: Dallas

CITY: LA JOLLA  
STATE: California  
COUNTY: SAN DIEGO

COUNTRY: US  
ZIP: 92037

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; GENERATED BY: TTY Control

```

```

;
;  COMPUTER:  IBM compatible
;  OPERATING SYSTEM:  DOS
;  COMMAND:  Test.CMD "word"

```

```

; SOFTWARE: FDSISEQ VERSION 1.3
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: 18-06-004 5000

```

APPLICATION NUMBER: 05/08/984, 709A  
FILING DATE: 02-DEC-1997

CLASSIFICATION: 433  
ATTORNEY/AGENT INFORMATION:

NAME: Seiamah, Stephanie L.  
REGISTRATION NUMBER: 33,779  
BIRTHDATE / DEATHDATE: 04/25/1981 - 03/03/2015

REFERENCE/DOCKET NUMBER: 24  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (610) 450-8400

TELEPHONE: (619) 430-8400  
TELEFAX: (619) 587-5360  
; INFORMATION FOR CTO TO NO. 8

```

; INFORMATION FOR SEQ ID NO: 8
;
; SEQUENCE CHARACTERISTICS:
;     LENGTH: 2175 base pairs

```

LENGTH: 1173 base pairs  
TYPE: nucleic acid  
COMMENT: double-stranded

TOPOLOGY: linear

```

;
; MOLECODE TYPE: DNA (genomic)
;
; FEATURE:
; NAME/KEY: COS

```

NAME/KEY: CDS  
LOCATION: 144..6857  
FEATURE:

```

FEATURE: 5'UTR
; NAME/KEY: 143
; LOCATION: 1

```

```

; LOCATION: 1..143
; FEATURE:
; NAME/KEY: 3/TTB

```

NAME/KEY: 5 01K  
LOCATION: 6855..7175  
; ;  
US-08-984-7092-8

05-06-204-103A-0

query match	32.18;	score 31.0;	DB 4;	length 11.3
Best Local Similarity	61.48;	Pred. No. 0.033;		

```
Matches 51; Conservative 0; Mismatches 32; Indels 0; Gaps 0;
OY 9 GAGCGCAAGCATGTGGCATGCTGTAAGACCACTAGATCCAGACGCCAGAC 68
    || ||||| | ||||| |||| || ||||| ||||
Db 1316 GATCTTAAGCGGAGAGATCATGCTGCCGAGAGACAGCATGCAGAGAGATC 1375
OY 69 CCCTTCTTTGTGCTCAAGAAAG 91
    ||||| ||| |||||
Db 1376 CCCTTGACGCTGCTGAAGAGAG 1398

RESULT 13
US-08-450-272-8
; Sequence 8, Application US/08450272
; Patent No. 6387696
; GENERAL INFORMATION:
; APPLICANT: Harpold, Michael
; APPLICANT: Ellis, Steven
; APPLICANT: Williams, Mark
; APPLICANT: McCue, Ann
; APPLICANT: Gillespie, Allison
; APPLICANT: Feldman, Daniel
; APPLICANT: Brenner, Robert
; TITLE OF INVENTION: HUMAN CALCIUM CHANNEL COMPOSITIONS AND
; TITLE OF INVENTION: METHODS
; NUMBER OF SEQUENCES: 38
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Brown, Martin, Haller & McClain
; STREET: 1660 Union Street
; CITY: San Diego
; STATE: California
; COUNTRY: US
; ZIP: 92101
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/450,272
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/404,950
; FILING DATE: 13-MAR-1995
; APPLICATION NUMBER: 08/336,257
; FILING DATE: 7-NOV-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/314,083
; FILING DATE: 28-SEPT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/311,363
; FILING DATE: 23-SEPT-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/290,012
; FILING DATE: 11-AUG-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/223,305
; FILING DATE: 4-APR-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/193,078
; FILING DATE: 07-FEB-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/149,097
; FILING DATE: 5-NOV-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/105,536
; FILING DATE: 11-AUG-1993
; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 07/914,231
; FILING DATE: 13-JULY-1992
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/868,354
; FILING DATE: 10-APR-1992
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/06903
; FILING DATE: 14-AUG-1992
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/745,206
; FILING DATE: 15-AUG-1991
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Seidman, Stephanie L.
; REGISTRATION NUMBER: 33,779
; REFERENCE/DOCKET NUMBER: 6362-519812
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 238-0999
; TELEFAX: (619) 238-0062
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7175 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 144..6857
; FEATURE:
; NAME/KEY: 5'UTR
; LOCATION: 1..143
; FEATURE:
; NAME/KEY: 3'UTR
; LOCATION: 6855..7175
; US-08-450-272-8

Query Match 32.1%; Score 31.8; DB 4; Length 7175;
Best Local Similarity 61.4%; Pred. No. 0.033;
Matches 51; Conservative 0; Mismatches 32; Indels 0; Gaps 0;
OY 9 GAGCGCAAGCATGTGGCATGCTGTAAGACCACTAGATCCAGACGCCAGAC 68
    || ||||| | ||||| |||| || ||||| ||||
Db 1316 GATCTTAAGCGGAGAGATCATGCTGCCGAGAGACAGCATGCAGAGAGATC 1375
OY 69 CCCTTCTTTGTGCTCAAGAAAG 91
    ||||| ||| |||||
Db 1376 CCCTTGACGCTGCTGAAGAGAG 1398

RESULT 14
US-09-268-163-7
; Sequence 7, Application US/09268163B
; Patent No. 6353091
; GENERAL INFORMATION:
; APPLICANT: Lipscombe, Diane
; APPLICANT: Schorge, Stephanie
; TITLE OF INVENTION: HUMAN N-TYPE CALCIUM CHANNEL ISOFORM AND USES THEREOF
; FILE REFERENCE: B1055/7000
; CURRENT APPLICATION NUMBER: US/09/268,163B
; CURRENT FILING DATE: 1999-03-12
; EARLIER APPLICATION NUMBER: US 60/077,901
; EARLIER FILING DATE: 1998-03-13
; NUMBER OF SEQ ID NOS: 28
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 7
; LENGTH: 7177
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
```

NAME/KEY: CDS  
LOCATION: 146..6856  
US-09-268-163-7

Query Match 32.1%; Score 31.8; DB 4; Length 7177;  
Best Local Similarity 61.4%; Pred. No. 0.033;  
Matches 51; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 9 GAGGCAAGGCAATGTCGTGTAAGACCACTAGATCCAGCAGCCAGAC 68  
DB 1318 GATCTTCAGGCGGAGGAGTCATGCTGCGGAGAGAGACAGATGCAAGAGAGAGTC 1377

QY 69 CCCTTTCTTTGTGTCAGAAAG 91  
DB 1378 CCCTTTGAGCGTCTGTAAGAGAG 1400

## RESULT 15

US-08-713-118-1  
Sequence 1, Application US/08713118  
Patent No. 6040436

## GENERAL INFORMATION:

APPLICANT: Franco, Rodrigo  
APPLICANT: Sun Chen, Ai Ru  
APPLICANT: Suey, David J.  
TITLE OF INVENTION: NUCLEIC ACID ENCODING HUMAN NEURONAL  
TITLE OF INVENTION: CALCIUM CHANNEL SUBUNITS  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.  
STREET: Two Militia Drive

CITY: Lexington  
STATE: MA

COUNTRY: USA  
ZIP: 02173-4799

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/713,118

FILING DATE: 16-SEP-1996

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Mata, Elizabeth W.

REGISTRATION NUMBER: 38,236

REFERENCE/DOCKET NUMBER: ACC96-01

TELECOMMUNICATION INFORMATION:

TELEPHONE: 617-861-6240

TELEFAX: 617-861-9540

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 7266 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: CDNA

FEATURE:

NAME/KEY: CDS  
LOCATION: 92..7102  
US-08-713-118-1

Query Match 32.1%; Score 31.8; DB 3; Length 7266;  
Best Local Similarity 61.4%; Pred. No. 0.033;  
Matches 51; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 9 GAGGCAAGGCAATGTCGTGTAAGACCACTAGATCCAGCAGCCAGAC 68  
DB 1264 GATCTTCAGGCGGAGGAGTCATGCTGCGGAGAGAGACAGATGCAAGAGAGAGTC 1323

QY 69 CCCTTTCTTTGTGTCAGAAAG 91  
||||| ||| | ||| | ||

DB 1324 CCCTTTGAGCGTCTGTAAGAGAG 1346

Search completed: May 2, 2003, 12:48:32  
Job time : 55.6 secs



GenCore version 5.1.5  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 2, 2003, 12:25:30 ; Search time 65,862 Seconds

(without alignments)  
1775.493 Million cell updates/sec

Title: US-09-719-554-3\_COPY\_2502\_2600

Perfect score: 99  
Sequence: 1 tcaccctgagcgcaagc.....tgcacgaagcgcgga 99

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 746064 seqs, 590810554 residues

Total number of hits satisfying chosen parameters: 1492128

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications\_MA:\*

- 1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq:\*
- 3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:\*
- 4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:\*
- 5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:\*
- 6: /cgn2\_6/ptodata/1/pubpna/PCTOS\_PUBCOMB.seq:\*
- 7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*
- 8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*
- 10: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq:\*
- 11: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:\*
- 12: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq:\*
- 13: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:\*
- 14: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	31.8	32.1	7177	12	US-10-033-026-7
2	31.8	32.1	7364	10	US-09-954-456-1179
3	31.8	32.1	7364	12	US-10-033-026-5
4	31.8	32.1	7376	12	US-10-033-026-3
5	30	30.3	7011	12	US-10-033-026-9
6	27.8	28.1	8794	9	US-09-764-891-8933
7	27	27.3	206	9	US-09-796-692-4050
8	27	27.3	206	9	US-10-040-862-4050
9	26.4	26.7	271	10	US-09-867-701-7757
10	26.4	26.7	6038	10	US-09-802-807-4
11	26.4	26.7	7622	10	US-09-802-807-1
12	26.2	26.5	424	9	US-09-934-531-665
13	26.2	26.5	14759	9	US-09-764-891-8344
14	26.2	26.5	30420	9	US-10-091-572-740
15	26.2	26.5	30420	9	US-09-764-891-9164
16	26.2	26.5	30420	10	US-09-764-877-2432
17	26	26.3	490	9	US-09-918-995-13251
18	26	26.3	1158	9	US-09-948-1993-2
19	26	26.3	3091	9	US-10-074-095-1001

20	26	26.3	3091	9	US-10-074-095-1002	Sequence 1002, Ap
21	26	26.3	3091	10	US-09-764-860-1001	Sequence 1001, Ap
22	26	26.3	3091	10	US-09-764-860-1002	Sequence 1002, Ap
23	25.8	26.1	167343	10	US-09-962-436-281	Sequence 281, App
24	25.8	26.1	167343	10	US-09-964-824A-273	Sequence 273, App
25	25.6	25.9	404	10	US-09-864-761-4275	Sequence 4275, Ap
26	25.6	25.9	479	9	US-09-918-995-9703	Sequence 9703, App
27	25.6	25.9	90541	10	US-09-759-359A-3	Sequence 3, Appl
28	25.4	25.7	245	10	US-09-878-574-13986	Sequence 13986, A
29	25.4	25.7	392	10	US-09-864-761-14440	Sequence 14440, A
30	25.4	25.7	759	10	US-09-867-274-1	Sequence 1, Appl
31	25.4	25.7	2329	9	US-10-245-103-63	Sequence 63, Appl
32	25.4	25.7	2329	9	US-10-245-107-63	Sequence 63, Appl
33	25.4	25.7	2329	9	US-10-245-143-63	Sequence 63, Appl
34	25.4	25.7	2329	9	US-10-245-771-63	Sequence 63, Appl
35	25.4	25.7	2329	9	US-10-245-851-63	Sequence 63, Appl
36	25.4	25.7	2329	9	US-10-245-883-63	Sequence 63, Appl
37	25.4	25.7	2329	9	US-10-237-535-63	Sequence 63, Appl
38	25.4	25.7	2329	9	US-10-238-183-63	Sequence 63, Appl
39	25.4	25.7	2329	9	US-10-238-183-63	Sequence 63, Appl
40	25.4	25.7	2329	9	US-10-238-370-63	Sequence 63, Appl
41	25.4	25.7	2329	9	US-10-245-055-63	Sequence 63, Appl
42	25.4	25.7	2329	9	US-10-245-147-63	Sequence 63, Appl
43	25.4	25.7	2329	9	US-10-245-730-63	Sequence 63, Appl
44	25.4	25.7	2329	9	US-10-245-739-63	Sequence 63, Appl
45	25.4	25.7	2329	9	US-10-246-210-63	Sequence 63, Appl

#### ALIGNMENTS

RESULT 1  
US-10-033-026-7  
; Sequence 7, Application US/10033026  
; Patent No. US20020147309A1  
; GENERAL INFORMATION:  
; APPLICANT: Schorger, Stephanie  
; TITLE OF INVENTION: HUMAN N-TYPE CALCIUM CHANNEL ISORFORM AND USES THEREOF  
; FILE REFERENCE: B1055/7000  
; CURRENT APPLICATION NUMBER: US/10/033, 026  
; PRIOR FILING DATE: 2001-12-28  
; PRIOR APPLICATION NUMBER: US 09/268, 163  
; PRIOR FILING DATE: 1999-03-12  
; PRIOR APPLICATION NUMBER: US 60/077, 901  
; NUMBER OF SEQ ID NOS: 28  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 7  
; LENGTH: 7177  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 146..6856  
US-10-033-026-7

Query Match 32.1%; Score 31.8; DB 12; Length 7177;  
Best Local Similarity 61.4%; Pred. No. 0.051;  
Matches 51; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 9 GACGCAAGACGATGTTGGCATGCGTGAAGACACATGACACAGCCAGAC 68  
DB 1318 GATCTTCACAGCGGAGAGACTGCTGCGCGGAGGAGACGAGATGACAGAGAGTC 1377

QY 69 CCCTTCTTGTGTGTCAGAAAG 91  
DB 1378 CCCTTGAAGCTGCTGAAGAG 1400

RESULT 2  
US-09-954-456-1179  
; Sequence 1179, Application US/09954456

```
Patent No. US20020115057A1
GENERAL INFORMATION:
APPLICANT: Young, Paul
TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Cand
TITLE OF INVENTION: Sets
FILE REFERENCE: 689290-76
CURRENT APPLICATION NUMBER: US/09/954,456
CURRENT FILING DATE: 2001-09-18
PRIOR APPLICATION NUMBER: US/60/233,617
PRIOR FILING DATE: 2000-09-18
PRIOR APPLICATION NUMBER: US/60/234,052
PRIOR FILING DATE: 2000-09-20
PRIOR APPLICATION NUMBER: US/60/234,923
PRIOR FILING DATE: 2000-09-25
PRIOR APPLICATION NUMBER: US/60/235,134
PRIOR FILING DATE: 2000-09-25
PRIOR APPLICATION NUMBER: US/60/235,637
PRIOR FILING DATE: 2000-09-26
PRIOR APPLICATION NUMBER: US/60/235,638
PRIOR FILING DATE: 2000-09-26
PRIOR APPLICATION NUMBER: US/60/235,711
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: US/60/235,720
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: US/60/235,840
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: US/60/235,863
PRIOR FILING DATE: 2000-09-27
NUMBER OF SEQ ID NOS: 2276
SOFTWARE: PatentIn version 3.0
SEQ ID NO 1179
LENGTH: 7364
TYPE: DNA
ORGANISM: Homo sapiens
US-09-954-456-1179

Query Match          32.1%; Score 31.8; DB 10; Length 7364;
Best Local Similarity 61.4%; Pred. No. 0.051;
Matches 51; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 9 GAGGCAAGGCAATGTTGGCATGCTGTAAAGACCACTAGAAATCCAGAGCCCGAC 68
DB 1318 GATCTTCAAGCGGAGAGATGATGCTGCGCCGAGAGACAGAAATCAGAGAGAGATC 1377
QY 69 CCCTTCTTTGTGTCAGAAAG 91
DB 1378 CCCTTTGAGCTGCTGTAAGAGAG 1400

RESULT 3
US-10-033-026-5
Sequence 5, Application US/10033026
Patent No. US20020147309A1
GENERAL INFORMATION:
APPLICANT: Lipscombe, Diane
APPLICANT: Schorge, Stephanie
TITLE OF INVENTION: HUMAN N-TYPE CALCIUM CHANNEL ISOFORM AND USES THEREOF
FILE REFERENCE: B1055/7000
CURRENT APPLICATION NUMBER: US/10/033,026
CURRENT FILING DATE: 2001-12-28
PRIOR APPLICATION NUMBER: US 09/268,163
PRIOR FILING DATE: 1999-03-12
PRIOR APPLICATION NUMBER: US 60/077,901
PRIOR FILING DATE: 1998-03-13
NUMBER OF SEQ ID NOS: 28
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 5
LENGTH: 7364
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: 146..7162
```

```
US-10-033-026-5
Query Match          32.1%; Score 31.8; DB 12; Length 7364;
Best Local Similarity 61.4%; Pred. No. 0.051;
Matches 51; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 9 GAGGCAAGGCAATGTTGGCATGCTGTAAAGACCACTAGAAATCCAGAGCCCGAC 68
DB 1318 GATCTTCAAGCGGAGAGATGATGCTGCGCCGAGAGACAGAAATCAGAGAGAGATC 1377
QY 69 CCCTTCTTTGTGTCAGAAAG 91
DB 1378 CCCTTTGAGCTGCTGTAAGAGAG 1400

RESULT 4
US-10-033-026-3
Sequence 3, Application US/10033026
Patent No. US20020147309A1
GENERAL INFORMATION:
APPLICANT: Lipscombe, Diane
APPLICANT: Schorge, Stephanie
TITLE OF INVENTION: HUMAN N-TYPE CALCIUM CHANNEL ISOFORM AND USES THEREOF
FILE REFERENCE: B1055/7000
CURRENT APPLICATION NUMBER: US/10/033,026
CURRENT FILING DATE: 2001-12-28
PRIOR APPLICATION NUMBER: US 09/268,163
PRIOR FILING DATE: 1999-03-12
PRIOR APPLICATION NUMBER: US 60/077,901
PRIOR FILING DATE: 1998-03-13
NUMBER OF SEQ ID NOS: 28
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 3
LENGTH: 7376
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: 146..7174
US-10-033-026-3

Query Match          32.1%; Score 31.8; DB 12; Length 7376;
Best Local Similarity 61.4%; Pred. No. 0.051;
Matches 51; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 9 GAGGCAAGGCAATGTTGGCATGCTGTAAAGACCACTAGAAATCCAGAGCCCGAC 68
DB 1318 GATCTTCAAGCGGAGAGATGATGCTGCGCCGAGAGACAGAAATCAGAGAGAGATC 1377
QY 69 CCCTTCTTTGTGTCAGAAAG 91
DB 1378 CCCTTTGAGCTGCTGTAAGAGAG 1400

RESULT 5
US-10-033-026-9
Sequence 9, Application US/10033026
Patent No. US20020147309A1
GENERAL INFORMATION:
APPLICANT: Lipscombe, Diane
APPLICANT: Schorge, Stephanie
TITLE OF INVENTION: HUMAN N-TYPE CALCIUM CHANNEL ISOFORM AND USES THEREOF
FILE REFERENCE: B1055/7000
CURRENT APPLICATION NUMBER: US/10/033,026
CURRENT FILING DATE: 2001-12-28
PRIOR APPLICATION NUMBER: US 09/268,163
PRIOR FILING DATE: 1999-03-12
PRIOR APPLICATION NUMBER: US 60/077,901
PRIOR FILING DATE: 1998-03-13
NUMBER OF SEQ ID NOS: 28
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 9
LENGTH: 7011
```

```

; TYPE: DNA
; ORGANISM: Rattus norvegicus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..7008
; US-10-033-026-9
```

```

Query Match
Best Local Similarity 30.3%; Score 30; DB 12; Length 7011;
Matches 51; Conservative 0; Mismatches 35; Indels 0; Gaps 0;
```

```

QY 9 GAGCGCAAGCAATGTTGGCATGTGTAAGACACACATCAATCCAGCCGAC 68
DB 1173 GATCTTCAAGGGGAGGAGTATGTGGCAGAGGACAGACGACGAGAGAGTC 1232
```

```

QY 69 CCCTTCTTTGTGTGTCAGAAAGCG 94
DB 1233 CCCTTGTGATGCAGTGTGAAGAGAG 1258
```

## RESULT 6

```

US-09-764-891-8933/c
; Sequence 8933, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8933
; LENGTH: 8794
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-764-891-8933
```

```

Query Match
Best Local Similarity 28.1%; Score 27.8; DB 9; Length 8794;
Matches 47; Conservative 0; Mismatches 32; Indels 0; Gaps 0;
```

```

QY 19 GCAATGTTGGCATGTGTAAGACACATCAATCCAGCCGACCCCTTCTTT 78
DB 4260 GCTATGTTGGCCACAGTGTGTTGAACTCTGCTCAAGCCTCCAAAGTGTGATTT 4201
```

```

QY 79 GTGTCAGAAAGGCGGGA 97
DB 4200 ATGGGCATGAGCGCGTGA 4182
```

## RESULT 7

```

US-09-796-692-4050/c
; Sequence 4050, Application US/09796692
; Publication No. US20020198362A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY
; FILE REFERENCE: 2077, 001200
; CURRENT APPLICATION NUMBER: US/09/796,692
; CURRENT FILING DATE: 2001-03-01
; PRIOR APPLICATION NUMBER: 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: 60/200,303
; PRIOR FILING DATE: 2000-04-28
```

```

; PRIOR APPLICATION NUMBER: 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: 60/218,950
; PRIOR FILING DATE: 2000-07-14
; PRIOR APPLICATION NUMBER: 60/222,903
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: 60/223,416
; PRIOR FILING DATE: 2000-08-04
; PRIOR APPLICATION NUMBER: 60/223,378
; PRIOR FILING DATE: 2000-08-07
; NUMBER OF SEQ ID NOS: 9597
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4050
; LENGTH: 206
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (77)
; OTHER INFORMATION: n=A,T,C or G
; US-09-796-692-4050
```

```

Query Match
Best Local Similarity 27.3%; Score 27; DB 9; Length 206;
Matches 51; Conservative 0; Mismatches 41; Indels 0; Gaps 0;
```

```

QY 1 TCACCCCTGAGCGCAAGGCATGTTGGCATGCTGTAAGGACCATAGATCCAGA 60
DB 160 TCAGCAGTGGCGCTGTAAGACAGGCTTGAGGATGATGAGCAGCATGCGCCAGAA 101
```

```

QY 61 GCCCAGACCCCTTCTTGTGTCAGAAAGG 92
DB 100 GCCCAGCTTCTCACCAGTGAAGGINTGAAGAGG 69
```

## RESULT 8

```

US-10-040-862-4050/c
; Sequence 4050, Application US/10040862
; Publication No. US20030078396A1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; APPLICANT: Retter, Marc
; APPLICANT: Corixa Corporation
; TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and The
; FILE REFERENCE: 014058-013520US
; CURRENT APPLICATION NUMBER: US/10/040,862
; CURRENT FILING DATE: 2001-11-06
; PRIOR APPLICATION NUMBER: US 60/186,126
; PRIOR FILING DATE: 2000-03-01
; PRIOR APPLICATION NUMBER: US 60/190,479
; PRIOR FILING DATE: 2000-03-17
; PRIOR APPLICATION NUMBER: US 60/200,545
; PRIOR FILING DATE: 2000-04-27
; PRIOR APPLICATION NUMBER: US 60/200,303
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,779
; PRIOR FILING DATE: 2000-04-28
; PRIOR APPLICATION NUMBER: US 60/200,999
; PRIOR FILING DATE: 2000-05-01
; PRIOR APPLICATION NUMBER: US 60/202,084
; PRIOR FILING DATE: 2000-05-04
; PRIOR APPLICATION NUMBER: US 60/206,201
; PRIOR FILING DATE: 2000-05-22
; PRIOR APPLICATION NUMBER: US 60/218,950
```

;; PRIOR FILING DATE: 2000-07-14  
;; PRIOR APPLICATION NUMBER: US 60/222,903  
;; PRIOR FILING DATE: 2000-08-03  
;; PRIOR APPLICATION NUMBER: US 60/223,416  
;; PRIOR FILING DATE: 2000-08-04  
;; PRIOR APPLICATION NUMBER: US 60/223,378  
;; PRIOR FILING DATE: 2000-08-07  
;; PRIOR APPLICATION NUMBER: US 09/796,692  
;; PRIOR FILING DATE: 2001-03-01  
;; NUMBER OF SEQ ID NOS: 10467  
;; SOFTWARE: FastSeq for Windows Version 3.0  
;; SEQ ID NO 4050  
;; LENGTH: 206  
;; TYPE: DNA  
;; ORGANISM: Homo sapiens  
;; FEATURE:  
;; NAME/KEY: unsure  
;; LOCATION: (77)  
;; OTHER INFORMATION: n=A,T,C or G  
US-10-040-862-4050

Query Match 27.3%; Score 27; DB 9; Length 206;  
Best Local Similarity 55.4%; Pred. No. 1.2;  
Matches 51; Conservative 0; Mismatches 41; Indels 0; Gaps 0;

QY 1 TCACCCCTGAGCGCAAGGCATGTTGGCATGCTGTAAGAGACCACAGATCGACA 60  
DB 160 TCAGAGAGGGGCGCTTAAGACAGAGGTTGGAGTGAGATGACAGACAGCCAGAA 101  
QY 61 GCCCAGACCCCTTCTTGTGTGTCAGAAAG 92  
DB 100 GCCCAGCTTCTCACAGTGTAGAGTGTGAGAG 69

RESULT 9  
US-09-867-701-7757  
;; Sequence 7757, Application US/09867701  
;; Patent No. US20020132237A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Aglate, Paul A.  
;; APPLICANT: Jones, Robert  
;; APPLICANT: Harlocker, Susan L.  
;; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
;; FILE REFERENCE: 210121.497  
;; CURRENT APPLICATION NUMBER: US/09/867,701  
;; CURRENT FILING DATE: 2001-05-29  
;; NUMBER OF SEQ ID NOS: 10912  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 7757  
;; LENGTH: 271  
;; TYPE: DNA  
;; ORGANISM: Homo sapiens  
US-09-867-701-7757

Query Match 26.7%; Score 26.4; DB 10; Length 271;  
Best Local Similarity 65.0%; Pred. No. 2.2;  
Matches 39; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 19 GCATGTTGGCATGCTGTAAGAGACCATAGATCCAGACGCCAGCCCTTCTTT 78  
DB 28 GCTATGTTGGCATGCTGTAAGAGACCATAGATCCAGACCAATCCCTGCTGCT 87

RESULT 10  
US-09-802-807-4/c  
;; Sequence 4, Application US/09802807  
;; Patent No. US20010034044A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Treco, Douglas A.  
;; APPLICANT: Seiden, Richard F.  
;; APPLICANT: Seiden, Richard F.  
;; TITLE OF INVENTION: GENOMIC SEQUENCES FOR PROTEIN PRODUCTION AND DELIVERY

;; FILE REFERENCE: 07236/016001  
;; CURRENT APPLICATION NUMBER: US/09/802,807  
;; CURRENT FILING DATE: 2001-03-08  
;; PRIOR APPLICATION NUMBER: 60/084,663  
;; PRIOR FILING DATE: 1998-05-07  
;; NUMBER OF SEQ ID NOS: 7  
;; SOFTWARE: FastSeq for Windows Version 3.0  
;; SEQ ID NO 4  
;; LENGTH: 6038  
;; TYPE: DNA  
;; ORGANISM: Homo sapiens  
US-09-802-807-4

Query Match 26.7%; Score 26.4; DB 10; Length 6038;  
Best Local Similarity 65.0%; Pred. No. 5.4;  
Matches 39; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 19 GCATGTTGGCATGCTGTAAGAGACCATAGATCCAGACGCCAGCCCTTCTTT 78  
DB 174 GCATGTTGGCATGCTGCTCAAAATCTAGACTCAAGCAATCCACCACTTGCTT 115

RESULT 11  
US-09-802-807-1/c  
;; Sequence 1, Application US/09802807  
;; Patent No. US20010034044A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Treco, Douglas A.  
;; APPLICANT: Seiden, Richard F.  
;; APPLICANT: Seiden, Richard F.  
;; TITLE OF INVENTION: GENOMIC SEQUENCES FOR PROTEIN PRODUCTION AND DELIVERY  
;; FILE REFERENCE: 07236/016001  
;; CURRENT APPLICATION NUMBER: US/09/802,807  
;; CURRENT FILING DATE: 2001-03-08  
;; PRIOR APPLICATION NUMBER: 60/084,663  
;; PRIOR FILING DATE: 1998-05-07  
;; NUMBER OF SEQ ID NOS: 7  
;; SOFTWARE: FastSeq for Windows Version 3.0  
;; SEQ ID NO 1  
;; LENGTH: 7622  
;; TYPE: DNA  
;; ORGANISM: Homo sapiens  
US-09-802-807-1

Query Match 26.7%; Score 26.4; DB 10; Length 7622;  
Best Local Similarity 65.0%; Pred. No. 5.8;  
Matches 39; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 19 GCATGTTGGCATGCTGTAAGAGACCATAGATCCAGACGCCAGCCCTTCTTT 78  
DB 174 GCATGTTGGCATGCTGCTCAAAATCTAGACTCAAGCAATCCACCACTTGCTT 115

RESULT 12  
US-09-954-531-665  
;; Sequence 665, Application US/09954531  
;; Patent No. US20020165180A1  
;; GENERAL INFORMATION:  
;; APPLICANT: Weaver, Zoe  
;; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using C  
;; FILE REFERENCE: 689290-77  
;; CURRENT APPLICATION NUMBER: US/09/954,531  
;; CURRENT FILING DATE: 2002-05-02  
;; PRIOR APPLICATION NUMBER: US/60/233,133  
;; PRIOR FILING DATE: 2000-09-18  
;; PRIOR APPLICATION NUMBER: US/60/234,009  
;; PRIOR FILING DATE: 2000-09-20  
;; PRIOR APPLICATION NUMBER: US/60/234,034  
;; PRIOR FILING DATE: 2000-09-20  
;; PRIOR APPLICATION NUMBER: US/60/234,509  
;; PRIOR FILING DATE: 2000-09-22  
;; PRIOR APPLICATION NUMBER: US/60/234,567



PRIOR FILING DATE: 2000-07-11	PRIOR APPLICATION NUMBER: 60/225,758
PRIOR FILING DATE: 2000-08-14	PRIOR APPLICATION NUMBER: 60/220,963
PRIOR FILING DATE: 2000-07-26	PRIOR APPLICATION NUMBER: 60/217,496
PRIOR FILING DATE: 2000-07-11	PRIOR APPLICATION NUMBER: 60/225,447
PRIOR FILING DATE: 2000-08-14	PRIOR APPLICATION NUMBER: 60/218,290
PRIOR FILING DATE: 2000-07-14	PRIOR APPLICATION NUMBER: 60/225,757
PRIOR FILING DATE: 2000-08-14	PRIOR APPLICATION NUMBER: 60/226,868
PRIOR FILING DATE: 2000-08-22	PRIOR APPLICATION NUMBER: 60/216,647
PRIOR FILING DATE: 2000-07-07	PRIOR APPLICATION NUMBER: 60/225,267
PRIOR FILING DATE: 2000-08-14	PRIOR APPLICATION NUMBER: 60/216,880
PRIOR FILING DATE: 2000-07-07	PRIOR APPLICATION NUMBER: 60/225,270
PRIOR FILING DATE: 2000-08-14	PRIOR APPLICATION NUMBER: 60/225,869
PRIOR FILING DATE: 2000-12-08	PRIOR APPLICATION NUMBER: 60/235,834
PRIOR FILING DATE: 2000-09-27	PRIOR APPLICATION NUMBER: 60/234,274
PRIOR FILING DATE: 2000-09-21	PRIOR APPLICATION NUMBER: 60/224,223
PRIOR FILING DATE: 2000-09-21	PRIOR APPLICATION NUMBER: 60/228,924
PRIOR FILING DATE: 2000-08-30	PRIOR APPLICATION NUMBER: 60/224,518
PRIOR FILING DATE: 2000-08-14	PRIOR APPLICATION NUMBER: 60/236,369
PRIOR FILING DATE: 2000-09-29	PRIOR APPLICATION NUMBER: 60/224,519
PRIOR FILING DATE: 2000-08-14	PRIOR APPLICATION NUMBER: 60/220,964
PRIOR FILING DATE: 2000-07-26	PRIOR APPLICATION NUMBER: 60/241,809
PRIOR FILING DATE: 2000-10-20	PRIOR APPLICATION NUMBER: 60/249,299
PRIOR FILING DATE: 2000-11-17	PRIOR APPLICATION NUMBER: 60/236,337
PRIOR FILING DATE: 2000-09-29	PRIOR APPLICATION NUMBER: 60/241,785
PRIOR FILING DATE: 2000-10-20	PRIOR APPLICATION NUMBER: 60/244,617
PRIOR FILING DATE: 2000-11-01	PRIOR APPLICATION NUMBER: 60/225,268
PRIOR FILING DATE: 2000-08-14	PRIOR APPLICATION NUMBER: 60/236,368
PRIOR FILING DATE: 2000-09-29	PRIOR APPLICATION NUMBER: 60/251,856
PRIOR FILING DATE: 2000-12-08	PRIOR APPLICATION NUMBER: 60/251,868
PRIOR FILING DATE: 2000-12-08	PRIOR APPLICATION NUMBER: 60/229,344
PRIOR FILING DATE: 2000-09-01	PRIOR APPLICATION NUMBER: 60/234,997
PRIOR FILING DATE: 2000-09-25	PRIOR APPLICATION NUMBER: 60/229,343
PRIOR FILING DATE: 2000-09-01	PRIOR APPLICATION NUMBER: 60/229,345
PRIOR FILING DATE: 2000-09-01	PRIOR APPLICATION NUMBER: 60/229,287
PRIOR FILING DATE: 2000-09-01	PRIOR APPLICATION NUMBER: 60/229,513
PRIOR FILING DATE: 2000-09-05	PRIOR APPLICATION NUMBER: 60/231,413
PRIOR FILING DATE: 2000-09-08	

Query Match	26.5%;	Score 26.2;	DB 9;	Length 30420;
Best Local Similarity	54.7%;	Pred. No. 10;		
Matches 52; Conservative	0;	Mismatches 43;	Indels 0;	Gaps 0;





GenCore version 5.1.5  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 2, 2003, 11:21:20 ; Search time 411.145 Seconds

(without alignments)  
6795.341 Million cell updates/sec

Title: US-09-719-554-3\_COPY\_3065\_3160

Perfect score: 96  
Sequence: 1 gccgcctgcacctcctgagg.....tgagatgaagtgcataagt 96

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl: \*  
1: gb.ba:\*  
2: gb.htg:\*  
3: gb.in:\*  
4: gb.om:\*  
5: gb.ov:\*  
6: gb.pat:\*  
7: gb.ph:\*  
8: gb.pl:\*  
9: gb.pr:\*  
10: gb.ro:\*  
11: gb.sts:\*  
12: gb.sy:\*  
13: gb.un:\*  
14: gb.vl:\*  
15: em.ba:\*  
16: em.fun:\*  
17: em.hum:\*  
18: em.in:\*  
19: em.mu:\*  
20: em.om:\*  
21: em.or:\*  
22: em.ov:\*  
23: em.pat:\*  
24: em.ph:\*  
25: em.pl:\*  
26: em.ro:\*  
27: em.sts:\*  
28: em.un:\*  
29: em.vl:\*  
30: em.htg.hum:\*  
31: em.htg.in:\*  
32: em.htg.other:\*  
33: em.htg.mus:\*  
34: em.htg.pln:\*  
35: em.htg.rtd:\*  
36: em.htg.mam:\*  
37: em.htg.vrt:\*  
38: em.sv:\*  
39: em.htgo.hum:\*  
40: em.htgo.mus:\*  
41: em.htgo.other:\*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	96	100.0	1321	6	AX000956	AX000956 Sequence
2	96	100.0	1321	6	AX027470	AX027470 Sequence
3	96	100.0	1321	6	AF072500	AF072500 Homo sapi
4	96	100.0	1326	6	AX007979	AX007979 Sequence
5	96	100.0	2372	6	AX000965	AX000965 Sequence
6	96	100.0	2372	6	AX027479	AX027479 Sequence
7	96	100.0	2372	6	AF072504	AF072504 Homo sapi
8	96	100.0	10499	6	AX007980	AX007980 Sequence
9	96	100.0	56093	6	AX329572	AX329572 Sequence
10	96	100.0	56093	6	HSAC000064	AC000064 Human BAC
11	96	100.0	149194	9	AC007566	AC007566 Homo sapi
12	95.6	99.6	7582	6	AX027480	AX027480 Sequence
13	95.6	98.3	2838	6	AX000957	AX000957 Sequence
14	94.4	98.3	2838	6	AX027471	AX027471 Sequence
15	94.4	98.3	2838	6	AF072499	AF072499 Homo sapi
16	94.4	98.3	2838	6	AF072499	AF072499 Homo sapi
17	71.4	74.4	139744	9	AL133513	AL133513 Human DNA
18	70.8	73.8	104853	9	AC117444	AC117444 Homo sapi
19	70.8	73.8	169462	2	AC024033	AC024033 Homo sapi
20	70.8	73.8	186723	2	AC027752	AC027752 Homo sapi
21	70.4	73.3	1774	6	AF045450	AF045450 Homo sapi
22	70.4	73.3	40205	9	AF045450	AF045450 Homo sapi
23	70.4	73.3	142742	9	AF121782	AF121782 Homo sapi
24	70.4	73.3	251124	9	HUAE000660	AE000660 Homo sapi
25	70.4	73.3	340000	9	HS21C080	AL163280 Homo sapi
26	69.8	72.7	94585	9	HS0401P4	AL109922 Human DNA
27	69.6	72.5	153803	2	AC055722	AC055722 Homo sapi
28	69.6	72.5	164310	2	AC016156	AC016156 Homo sapi
29	69.6	72.5	182224	2	AC093116	AC093116 Homo sapi
30	69.6	72.5	203330	2	AC026346	AC026346 Homo sapi
31	69.2	72.1	70352	9	AL592310	AL592310 Human DNA
32	69.2	72.1	25437	9	AP000500	AP000500 Homo sapi
33	69.2	72.1	270000	9	AB026898	AB026898 Homo sapi
34	68.8	71.7	66369	2	AC090325	AC090325 Homo sapi
35	68.8	71.7	101196	9	AC087436	AC087436 Homo sapi
36	68.8	71.7	136901	9	AC073626	AC073626 Homo sapi
37	68.8	71.7	147655	9	AC104716	AC104716 Homo sapi
38	68.8	71.7	150955	9	AL356111	AL356111 Human DNA
39	68.8	71.7	155925	9	AC096536	AC096536 Homo sapi
40	68.8	71.7	157043	2	AC013692	AC013692 Homo sapi
41	68.8	71.7	161571	2	AC092708	AC092708 Homo sapi
42	68.8	71.7	161938	9	AL360227	AL360227 Human DNA
43	68.8	71.7	166107	2	AC024427	AC024427 Homo sapi
44	68.8	71.7	166700	9	AC013759	AC013759 Homo sapi
45	68.8	71.7	169072	9	AC009946	AC009946 Homo sapi

#### ALIGNMENTS

RESULT 1  
AX000956 1321 bp DNA linear PAT 10-MAR-2000  
LOCUS AX000956  
DEFINITION Sequence 1 from Patent WO9902656.  
ACCESSION AX000956  
VERSION AX000956.1 GI:7241198  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE 1 (bases 1 to 1321)  
AUTHORS Beseme, F. and Blond, J.  
TITLE ENDOGENETIC RETROVIRAL SEQUENCES, ASSOCIATED WITH AUTOIMMUNE  
DISEASES OR WITH PREGNANCY DISORDERS  
JOURNAL Patent: WO 9902656-A 1 21-JAN-1999;

Pred. No. is the number of results predicted by chance to have a

FEATURES BIO MERIEUX (FR); BESEME FREDERIC (FR)  
source location/Qualifiers  
1. .1321  
/organism="unidentified"  
/db\_xref="taxon:32644"  
BASE COUNT 365 a 352 c 277 g 327 t  
ORIGIN

Query Match 100.0%; Score 96; DB 6; Length 1321;  
Best Local Similarity 100.0%; Pred. No. 7.9e-19;  
Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCGCTGGACCTCTGAGGAGATATAATATACACCATCTTACAGCTAGACCTCTT 60  
|||||  
DB 717 GCCGCTGGACCTCTGAGGAGATATAATATACACCATCTTACAGCTAGACCTCTT 776  
|||||

QY 61 TTGTAGAAAAGCAATGAGTGAAGTCCCTAAGT 96  
|||||  
DB 777 TTGTAGAAAAGCAATGAGTGAAGTCCCTAAGT 812  
|||||

RESULT 2  
AX027470 1321 bp DNA linear PAT 16-SEP-2000  
LOCUS Sequence 20 from Patent FR2788784.  
DEFINITION AX027470  
ACCESSION AX027470.1 GI:10188434  
VERSION  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheraia; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 1321)  
AUTHORS Mallet,F., Voisset,C. and Paranhos,B.G.  
JOURNAL Patent: FR 2788784-A 20 28-JUL-2000;  
BIO MERIEUX (FR)

FEATURES location/Qualifiers  
source 1. .1321  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
BASE COUNT 365 a 352 c 277 g 327 t  
ORIGIN

Query Match 100.0%; Score 96; DB 6; Length 1321;  
Best Local Similarity 100.0%; Pred. No. 7.9e-19;  
Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCGCTGGACCTCTGAGGAGATATAATATACACCATCTTACAGCTAGACCTCTT 60  
|||||  
DB 717 GCCGCTGGACCTCTGAGGAGATATAATATACACCATCTTACAGCTAGACCTCTT 776  
|||||

QY 61 TTGTAGAAAAGCAATGAGTGAAGTCCCTAAGT 96  
|||||  
DB 777 TTGTAGAAAAGCAATGAGTGAAGTCCCTAAGT 812  
|||||

RESULT 3  
AF072500 1321 bp mRNA linear PRI 10-FEB-1999  
LOCUS AF072500 Homo sapiens endogenous retrovirus W sequence.  
DEFINITION AF072500  
ACCESSION AF072500.1 GI:4262284  
VERSION  
KEYWORDS  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheraia; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 1321)  
AUTHORS Blond,J.L., Beseme,F., Durel,L., Bouton,O., Bedin,F., Perron,H.,  
Mandrand,B. and Mallet,F.  
JOURNAL Molecular characterization and placental expression of HENV-W, a  
new human endogenous retrovirus family  
J. VIROL. 73 (2), 1175-1185 (1999)

MEDLINE 99099005  
PUBMED 9882319  
REFERENCE 2 (bases 1 to 1321)  
AUTHORS Blond,J.L., Beseme,F. and Mallet,F.  
TITLE Direct Submission  
JOURNAL Submitted (19-JUN-1998) UM103 CNRS-BioMerieux, ENS Lyon, 46 allée  
d'Italie, Lyon, Cedex 07 69364, France

FEATURES location/Qualifiers  
source 1. .1321  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="cl.6a2"  
/tissue\_type="placenta"  
/clone\_1b="Clontech 5'-stretch plus library, Cat number  
HLS014a"  
misc\_feature 1. .1321  
/note="corresponds to gag"  
repeat\_region 1. .1321  
/rpt\_family="HERV-W"  
5'UTR 1. .120  
/note="5' end of R uncertain"  
5'UTR 121..574  
/note="05"  
578..595  
primer\_bind /note="putative primer binding site for tRNA-W"  
BASE COUNT 365 a 352 c 277 g 327 t  
ORIGIN

Query Match 100.0%; Score 96; DB 9; Length 1321;  
Best Local Similarity 100.0%; Pred. No. 7.9e-19;  
Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCGCTGGACCTCTGAGGAGATATAATATACACCATCTTACAGCTAGACCTCTT 60  
|||||  
DB 717 GCCGCTGGACCTCTGAGGAGATATAATATACACCATCTTACAGCTAGACCTCTT 776  
|||||

QY 61 TTGTAGAAAAGCAATGAGTGAAGTCCCTAAGT 96  
|||||  
DB 777 TTGTAGAAAAGCAATGAGTGAAGTCCCTAAGT 812  
|||||

RESULT 4  
AX007979 1326 bp DNA linear PAT 06-SEP-2000  
LOCUS AX007979  
DEFINITION Sequence 2 from Patent WO967395.  
ACCESSION AX007979  
VERSION AX007979.1 GI:9995676  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheraia; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 1326)  
AUTHORS Perrin,J.P., Rieger,F. and Alliel,P.M.  
TITLE Nucleic sequence and deduced protein sequence family with human  
endogenous retroviral motifs, and their uses  
JOURNAL Patent: WO 967395-A 2 29-DEC-1999;  
INST NAT SANTE RECH MED (FR); PERIN JEAN PIERRE (FR); RIEGPR  
FRANCOIS (FR); ALLIEL PATRICK M (FR)

FEATURES location/Qualifiers  
source 1. .1326  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
BASE COUNT 425 a 306 c 303 g 292 t  
ORIGIN

Query Match 100.0%; Score 96; DB 6; Length 1326;  
Best Local Similarity 100.0%; Pred. No. 7.9e-19;  
Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCGCTGGACCTCTGAGGAGATATAATATACACCATCTTACAGCTAGACCTCTT 60  
|||||  
DB 1 GCCGCTGGACCTCTGAGGAGATATAATATACACCATCTTACAGCTAGACCTCTT 60  
|||||

QY 61 TTGTAGAAAAGCAATGAGTGAAGTGCCTAAGT 96  
|||||  
Db 61 TTGTAGAAAAGCAATGAGTGAAGTGCCTAAGT 96

RESULT 5  
LOCUS AX000965 2372 bp DNA linear PAT 10-MAR-2000  
DEFINITION Sequence 10 from Patent WO9902696.  
ACCESSION AX000965  
VERSION AX000965.1 GI:7241207  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified.

REFERENCE  
1 (bases 1 to 2372)  
Beseme, F. and Blond, J.  
ENDGENETIC RETROVIRAL SEQUENCES, ASSOCIATED WITH AUTOIMMUNE  
DISEASES OR WITH PREGNANCY DISORDERS  
Patent: WO 9902696-A-10 21-JAN-1999;  
JOURNAL BIO MERIEUX (FR); BESEME FREDERIC (FR)  
LOCATION/Qualifiers  
1. 2372  
/organism="unidentified"  
/db\_xref="taxon:32644"

BASE COUNT 674 a 593 c 544 g 557 t 4 others  
ORIGIN

Query Match 100.0%; Score 96; DB 6; Length 2372;  
Best Local Similarity 100.0%; Pred. No. 7.2e-19;  
Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCGCTGGCAGCTCTGAGGGAAGTAAATTAACACCATCTTACAGCTAGACCTCTT 60  
|||||  
Db 603 GCCGCTGGCAGCTCTGAGGGAAGTAAATTAACACCATCTTACAGCTAGACCTCTT 662

QY 61 TTGTAGAAAAGCAATGAGTGAAGTGCCTAAGT 96  
|||||  
Db 663 TTGTAGAAAAGCAATGAGTGAAGTGCCTAAGT 698

RESULT 6  
LOCUS AX027479 2372 bp DNA linear PAT 16-SEP-2000  
DEFINITION Sequence 29 from Patent FR278784.  
ACCESSION AX027479  
VERSION AX027479.1 GI:10188443  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE  
1 (bases 1 to 2372)  
Mallet, F., Voisset, C. and Paranhos, B. G.  
Patent: FR 2788784-A 29 28-JUL-2000;  
JOURNAL BIO MERIEUX (FR)  
LOCATION/Qualifiers  
1. 2372  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"

BASE COUNT 674 a 593 c 544 g 557 t 4 others  
ORIGIN

Query Match 100.0%; Score 96; DB 6; Length 2372;  
Best Local Similarity 100.0%; Pred. No. 7.2e-19;  
Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCGCTGGCAGCTCTGAGGGAAGTAAATTAACACCATCTTACAGCTAGACCTCTT 60  
|||||  
Db 603 GCCGCTGGCAGCTCTGAGGGAAGTAAATTAACACCATCTTACAGCTAGACCTCTT 662

QY 61 TTGTAGAAAAGCAATGAGTGAAGTGCCTAAGT 96  
|||||

Db 663 TTGTAGAAAAGCAATGAGTGAAGTGCCTAAGT 698  
|||||

RESULT 7  
LOCUS AF072504 2372 bp mRNA linear PRI 10-FEB-1999  
DEFINITION Homo sapiens endogenous retrovirus W sequence.  
ACCESSION AF072504  
VERSION AF072504.1 GI:4262288  
KEYWORDS  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE  
1 (bases 1 to 2372)  
Blond, J. L., Beseme, F., Duret, L., Boulton, O., Bedin, F., Perron, H.,  
Mandrard, B. and Mallet, F.  
Molecular characterization and placental expression of HERV-W, a  
new human endogenous retrovirus family  
J. Virol. 73 (2), 1175-1185 (1999)

JOURNAL  
MEDLINE  
PUBMED  
99099005  
9882319

REFERENCE  
2 (bases 1 to 2372)  
Blond, J. L., Beseme, F. and Mallet, F.  
Direct Submission  
Submitted (19-JUN-1998) UM103 CNRS-biomerieux, ENS Lyon, 46 allée  
d'Italie, Lyon, Cedex 07 69364, France  
LOCATION/Qualifiers  
1. 2372  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="Cl.44.4"  
/tissue\_type="placenta"  
/clone\_lib="Clontech 5'-stretch plus library, Cat number  
HL5014a"  
1. 2372  
/note="corresponds to gag and pol"

misc\_feature  
repeat\_region  
5' UTR  
5' UTR  
5' UTR  
primer\_bind  
BASE COUNT 674 a 593 c 544 g 557 t 4 others  
ORIGIN

Query Match 100.0%; Score 96; DB 9; Length 2372;  
Best Local Similarity 100.0%; Pred. No. 7.2e-19;  
Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCGCTGGCAGCTCTGAGGGAAGTAAATTAACACCATCTTACAGCTAGACCTCTT 60  
|||||  
Db 603 GCCGCTGGCAGCTCTGAGGGAAGTAAATTAACACCATCTTACAGCTAGACCTCTT 662

QY 61 TTGTAGAAAAGCAATGAGTGAAGTGCCTAAGT 96  
|||||  
Db 663 TTGTAGAAAAGCAATGAGTGAAGTGCCTAAGT 698

RESULT 8  
LOCUS AX007980 10499 bp DNA linear PAT 06-SEP-2000  
DEFINITION Sequence 3 from Patent WO9967395.  
ACCESSION AX007980  
VERSION AX007980.1 GI:9995677  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 10499)  
AUTHORS Perlin,J.P., Rieger,F. and Alliel,P.M.  
TITLE Nucleic sequence and deduced protein sequence family with human  
JOURNAL INST NAT SANTE RECH MED (FR); PERIN JEAN PIERRE (FR); RIEGER  
FRANCOIS (FR); ALLIEL PATRICK M (FR)  
FEATURES  
source 1. 10499  
/db\_xref="taxon:9606"  
BASE COUNT 3048 a 2676 c 2280 g 2495 t  
ORIGIN

Query Match 100.0%; Score 96; DB 6; Length 10499;  
Best Local Similarity 100.0%; Pred. No. 5.5e-19;  
Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCGCGCTGGCAGCTCTGAGGAGATTAATTAACACCATCTTACAGCTAGCTCTT 60  
|||||  
Db 3065 GCCGCGCTGGCAGCTCTGAGGAGATTAATTAACACCATCTTACAGCTAGCTCTT 3124  
|||||

OY 61 TTGTAGAAAAGCAATGAGTGAAGTGCATAGT 96  
|||||  
Db 3125 TTGTAGAAAAGCAATGAGTGAAGTGCATAGT 3160  
|||||

RESULT 9  
AX329572 56093 bp DNA linear PAT 09-JAN-2002  
LOCUS  
DEFINITION Sequence 81 from Patent WO0194629.  
ACCESSION AX329572  
VERSION AX329572.1 GI:18102550  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1  
Young,P.E., Augustus,M., Carter,K.C., Ebner,R., Endress,G.,  
Horrigan,S., Soppet,D.R. and Weaver,Z.  
TITLE Cancer gene determination and therapeutic screening using signature  
JOURNAL Patent: WO 0194629-A 81 13-DEC-2001;  
FEATURES  
source 1. 56093  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
BASE COUNT 16164 a 12346 c 10702 g 16881 t  
ORIGIN

Query Match 100.0%; Score 96; DB 6; Length 56093;  
Best Local Similarity 100.0%; Pred. No. 4.1e-19;  
Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCGCGCTGGCAGCTCTGAGGAGATTAATTAACACCATCTTACAGCTAGCTCTT 60  
|||||  
Db 31065 GCCGCGCTGGCAGCTCTGAGGAGATTAATTAACACCATCTTACAGCTAGCTCTT 31124  
|||||

OY 61 TTGTAGAAAAGCAATGAGTGAAGTGCATAGT 96  
|||||  
Db 31125 TTGTAGAAAAGCAATGAGTGAAGTGCATAGT 31160  
|||||

RESULT 10  
HSAC000064 56093 bp DNA linear PRI 13-NOV-1996  
LOCUS  
DEFINITION Human BAC clone RG083M05 from 7q21-7q22, complete sequence.  
ACCESSION AC000064  
VERSION AC000064.1 GI:1669369  
KEYWORDS HTG.  
SOURCE Homo sapiens.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 56093)  
AUTHORS Pauley,A.  
TITLE The sequence of H. sapiens BAC clone RG083M05  
JOURNAL Unpublished (1996)  
REFERENCE 2 (bases 1 to 56093)  
AUTHORS Waterston,R.  
TITLE Direct Submission  
JOURNAL Submitted (13-NOV-1996)  
COMMENT Genome Sequencing Center  
Department of Genetics, Washington University  
St. Louis, MO 63108, USA  
e-mail: sapiens@watson.wustl.edu

NOTICE: This sequence may not represent the entire insert of this  
clone. It may be shorter because we only sequence overlapping  
sections once, or longer because we provide a small overlap between  
neighboring submissions.

This sequence was finished as follows unless otherwise noted:  
all regions were double stranded or sequenced with an alternate  
chemistry; an attempt was made to resolve all sequencing problems,  
such as compressions and repeats; all regions were covered by  
sequence from more than one subclone; and the assembly was  
confirmed by restriction digest.

SOURCE INFORMATION:  
This clone is from the first release of the human BAC library. The  
library contains cloned DNA from a human male fibroblast cell line  
978SK. For references see: Shizuya et al., Proc. Natl. Acad. Sci.  
89:8794-8797 (1992); Kim et al., Genomics 34:213-218 (1996).  
VECTOR: pBEO  
Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:  
The orientation of this clone is unknown. Actual start of this  
clone is at base position 1 of H\_RG083M05, actual end is at 56093  
of H\_RG083M05

FEATURES  
source 1. 56093  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="7"  
/map="7q21-7q22"  
/clone="H\_RG083M05"  
/clone\_lib="CITB-978SK-B"  
complement(838..1131)  
/rpt\_family="ALU"  
<360..16971  
/gene="WUSC:H\_RG083M05.1"  
join(<1360..1503,4181..4370,4587..4774,6422..6556,  
9483..9547,11631..11773,11864..12021,13131..13296,  
14885..14988,16349..16546,16837..16971)  
/gene="WUSC:H\_RG083M05.1"  
/note="ATPase, strong similarity to peroxisome  
biogenesis protein PAB1 (PID:91172019); coded for by  
human cDNA C04279 (NID:91467530)"  
/codon\_start=1  
/protein\_id="AAB46346.1"  
/db\_xref="GI:1669371"  
/translation="KRLENIQKLEVAESEAVMOPSVVLDDDLINGLPAVPEHEH  
SPDAERCEILCNVINKLDCDINKFTDLDLHVAKEGTGFVADFTVLVDAISHRL  
SRQISIRKELVLTLDLFOKALRGFLPALSRSVNLHKPRDLGMDKIKGHEVROIIMD  
TIOLPKAVCLKREKYEPELPANLPIRGRTGILYGPCTGTLGAGVIAESMNFISV  
KQBELSKYIGASEQAVRDIIFRAQAKCILEPFEESIAIRGRDNGVMDRVNVO  
LITOLDGVGICGVYLAATSRPDLIPALLRPRDKVCVCPDPOVYVSYLESKTO  
OMHSFLVRLLEINLVSDSLPLADDVDLQHYASTVDSFTGDLKALALYNAOLEALHG  
MLSKSEILPDESKFNMYRFLYFGSEISLNGTSSDLSOCLAPSSMTDGLGCPV  
GDOLEFSQPPVLRASQEGCELTQEQRLRADISIIIGRYRSOSGEDESNMGPI



KTR1A1S0SHLMTALGHTRPISSEDWKNFAEL"

```

repeat_region complement(4948..5130)
                /rpt_family="ALU"
repeat_region complement(6581..7133)
                /rpt_family="L1"
repeat_region complement(7767..8037)
                /rpt_family="ALU"
repeat_region complement(8186..8472)
                /rpt_family="ALU"
misc_feature 8473..8625
              /note="WUGSC:H.RG083M05.1"
              /note="match to human 3' EST H75782 (NID:g1049794), bases
              287-444"
misc_feature 8841..9161
              /gene="WUGSC:H.RG083M05.1"
              /note="match to human 5' EST H75921 (NID:g1050050), bases
              21-348"
              /note="match to human 5' EST H75921 (NID:g1050050), bases
              9481..9547"
              /gene="WUGSC:H.RG083M05.1"
              /note="match to human 5' EST N22627 (NID:g1130501), bases
              276-343"
repeat_region complement(12612..12907)
                /rpt_family="ALU"
misc_feature 13670..13793
              /gene="WUGSC:H.RG083M05.1"
              /note="match to human 5' EST H41382 (NID:917434), bases
              143-266"
              /note="match to human 5' EST H41382 (NID:917434), bases
              13794..13877"
              /rpt_family="ALU"
              /rpt_family="L1"
              /gene="WUGSC:H.RG083M05.1"
              /note="match to human 5' EST H41382 (NID:917434), bases
              30-58"
              /note="match to human 5' EST H41382 (NID:917434), bases
              13907..14104"
              /rpt_family="ALU"
repeat_region complement(14110..14137)
                /rpt_family="L1"
repeat_region complement(15618..15907)
                /rpt_family="ALU"
              17227..17522
              /rpt_family="ALU"
              18667..19235
              /note="match to human fetal brain 5' EST D61494
              (NID:g970409), bases 1-255, and to human 3' EST R07476
              (NID:g759399)"
              19550..19670
              /rpt_family="ALU"
              21507..37303
              /note="similarity to various ss-RNA virus polypeptides;
              pseudogene: region of matches and close matches to
              multiple human ESTs, see R68740 (NID:g842257)"
              37316..37489
              /note="Grail prediction, score = 80"
              /evidence=not_experimental
              complement(38938..39224)
              /rpt_family="ALU"
              39225..39707
              /note="match to multiple human ESTs, see N30113
              (NID:g1148633)"
              39800..40085
              /rpt_family="ALU"
              complement(40247..40538)
              /rpt_family="ALU"
              complement(40632..40924)
              /rpt_family="ALU"
              complement(42283..42891)
              /rpt_family="ALU"
              complement(45474..45613)
              /rpt_family="ALU"
              complement(45614..45737)
              /note="match to human 3' EST H48898 (NID:g988738), bases
              129-333"
              complement(46107..47026)

```

```

repeat_region complement(47027..47318)
                /rpt_family="ALU"
misc_feature complement(47365..47782)
              /note="match to multiple human ESTs, see W37495
              (NID:g1319089)"
              47898..48115
              /note="match to human 5' EST H62306 (NID:g1015138), bases
              93-368"
repeat_region complement(48116..48405)
                /rpt_family="ALU"
misc_feature complement(48406..48584)
              /note="match to human 3' EST N29952 (NID:g1148472), bases
              290-455, and 5' EST R12730 (NID:g765806)"
              complement(48787..49405)
              /rpt_family="ALU"
              complement(49406..49534)
              /note="match to human 3' EST R65794 (NID:g838432), bases
              309-440"
              complement(49638..49672)
              /rpt_family="ALU"
              complement(49674..49890)
              /note="match to human 3' EST N29952 (NID:g1148472) and 5'
              EST N29938 (NID:g1148458), sequences are from opposite
              ends of the same clone"
              complement(49698..51806)
              /gene="WUGSC:H.RG083M05.2"
              complement(join(49698..49888,51575..51806))
              /gene="WUGSC:H.RG083M05.2"
              /note="WUGSC:H.RG083M05.2"
              /note="coded for by human cDNAs W37389 (NID:g1139205),
              R65891 (NID:g838529), R65794 (NID:g838432) and R65794
              (NID:g838432)"
              /codon_start=1
              /protein_id="AAB46345.1"
              /db_xref="GI:1669370"
              /translation="MEFYRQCGIIFRPGYVYQIDVYSVIDEGRKYAQRGF
              IODYCKSAALTWLPTLSSPRQDFPASTIIQFEDLPKMYELFVCHAPSEYRK
              SRSSPFPTVPTRPEKGYTWTHVGEPPIITIKESVANH"
              complement(51576..51758)
              /gene="WUGSC:H.RG083M05.2"
              /note="Grail prediction, score = 86"
              /evidence=not_experimental
              complement(52052..52325)
              /rpt_family="L1"
              55557..55843
              /note="match to human EST M79192 (NID:g273505) base 2-289"

```

Query Match 100.0%; Score 96; DB 9; Length 56093;  
 Best Local Similarity 100.0%; Pred. No. 4.1e-19;  
 Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 GCGCGTCGCACTCCTGAGGGAAGATATAATTTATACACATCTTACAGTAGACCTCTT 60
    |||||
Db 31065 GCGCGTCGCACTCCTGAGGGAAGATATAATTTATACACATCTTACAGTAGACCTCTT 31124
    |||||
QY 61 TTGTAGAAAGCAAAATGAGTGAAGTCCATAACT 96
    |||||
Db 31125 TTGTAGAAAGCAAAATGAGTGAAGTCCATAACT 31160
    |||||

```

RESULT 11  
 AC007566/c 149194 bp DNA linear PRI 01-MAR-2002  
 LOCUS Homo sapiens BAC clone CTB-1065 from 7q21-7q22, complete sequence.  
 DEFINITION AC007566  
 ACCESSION AC007566.2 GI:11181861  
 VERSION  
 KEYWORDS HIG  
 SOURCE Homo sapiens.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE 1 (bases 1 to 149194)  
 AUTHORS Sulston,J.E. and Waterston,R.  
 TITLE Toward a complete human genome sequence  
 JOURNAL Genome Res. 8 (11), 1097-1108 (1998)  
 MEDLINE 99063792  
 PUBMED 9847074  
 REFERENCE 2 (bases 1 to 149194)  
 AUTHORS Du,Z.  
 TITLE The sequence of Homo sapiens BAC clone CTB-10G5  
 JOURNAL Unpublished (2001)  
 REFERENCE 3 (bases 1 to 149194)  
 AUTHORS Waterston,R.H.  
 TITLE Direct Submission  
 JOURNAL Submitted (15-MAY-1999) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA  
 REFERENCE 4 (bases 1 to 149194)  
 AUTHORS Waterston,R.  
 TITLE Direct Submission  
 JOURNAL Submitted (02-OCT-2000) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
 REFERENCE 5 (bases 1 to 149194)  
 AUTHORS Waterston,R.H.  
 TITLE Direct Submission  
 JOURNAL Submitted (16-NOV-2000) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA  
 REFERENCE 6 (bases 1 to 149194)  
 AUTHORS Waterston,R.H.  
 TITLE Direct Submission  
 JOURNAL Submitted (03-JAN-2002) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA  
 REFERENCE 7 (bases 1 to 149194)  
 AUTHORS Waterston,R.H.  
 TITLE Direct Submission  
 JOURNAL Submitted (06-FEB-2002) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA  
 REFERENCE 8 (bases 1 to 149194)  
 AUTHORS Waterston,R.  
 TITLE Direct Submission  
 JOURNAL Submitted (01-MAR-2002) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
 On Nov 16, 2000 this sequence version replaced gi:4835815.  
 ----- Genome Center  
 Center: Washington University Genome Sequencing Center  
 Center code: WUGSC  
 Web site: <http://genome.wustl.edu/gsc>  
 Contact: [saplens@wustl.edu](mailto:saplens@wustl.edu)  
 ----- Summary Statistics  
 Center project name: H\_RG010G05

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:  
 all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:  
 The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information

about the map position of this sequence, see  
<http://www.nhgri.nih.gov/DIR/CTB/CHR7>, send  
<mailto:egreen@nhgri.nih.gov>, or see <http://genome.wustl.edu/gsc>

#### SOURCE INFORMATION:

Clone CTB-10G5 is from the first release of the human BAC library CTB-978SK-B. The library contains cloned DNA from the male fibroblast cell line 978SK. See: Shizuya et al., Proc. Natl. Sci. USA 89:8794-7 (1992); U-J. Kim et al., Genomics 34:213-8 (1996). This clone is available from Research Genetics, Inc. (<http://www.resgen.com>).  
 VECTOR: pGelBAC11  
 Selection: chloramphenicol

#### NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP5-911H5, 200 base pair overlap. Actual start of this clone is at base position 195 of CTB-10G5; actual end is at base position 150532 of CTB-10G5.

#### FEATURES

##### source

The clone CTB-10G5 contains the entire sequence of CTB-83M5.  
 Location/Qualifiers

```

1..149194
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/map="7q21-7q22"
/clone="CTB-10G5"
/clone_11b="CTB-978SK-B"

1..1634
/rpt_family="L1"
misc_feature
2248..2388
/note="match to EST BG752883 (NID:g14063536)"
misc_feature
2248..2387
/note="match to EST AA149693 (NID:g1720635) zn99406.r1"
misc_feature
2248..2387
/note="match to EST AM579261 (NID:g7254310)"
misc_feature
2248..2387
/note="match to EST BG766882 (NID:g14077535)"
misc_feature
2248..2387
/note="match to EST B1160365 (NID:g14620366)"
misc_feature
2248..2374
/note="similar to Homo sapiens EST BF758865 (NID:g12106765)"
misc_feature
2248..2287
/note="match to EST AV686676 (NID:g10288539)"
misc_feature
2253..2387
/note="match to EST AU123510 (NID:g10948226)"
misc_feature
2344..2387
/note="match to EST AV686676 (NID:g10288539)"
misc_feature
2696..3066
/rpt_family="MaLR"
repeat_region
3108..3392
/rpt_family="Alu"
misc_feature
3540..3628
/note="match to EST AA425526 (NID:g2106267) zw48b03.r1"
misc_feature
3540..3628
/note="similar to Mus musculus EST BB253526 (NID:g8946272)"
misc_feature
3542..3632
/note="match to EST AV686676 (NID:g10288539)"
misc_feature
3542..3628
/note="match to EST AA149693 (NID:g1720635) zn99406.r1"
misc_feature
3542..3628
/note="match to EST AU123510 (NID:g10948226)"
misc_feature
3542..3628
/note="match to EST BG766882 (NID:g14077535)"
misc_feature
3542..3628
/note="match to EST B1160365 (NID:g14620366)"
misc_feature
3542..3609
/note="match to EST AM579261 (NID:g7254310)"
misc_feature
3562..3628
/note="match to EST BE272564 (NID:g9146913)"
misc_feature
3598..3628
```

misc\_feature /note="match to EST BE299708 (NID:g9183456)"  
3714..3785  
/note="match to EST BG260659 (NID:g12770475)"  
3717..3785  
misc\_feature /note="similar to Mus musculus EST BE994936  
(NID:g10678674)"  
3719..3785  
misc\_feature /note="match to EST AA149693 (NID:g120635) zn99d06.r1"  
3719..3785  
misc\_feature /note="match to EST AA425526 (NID:g2106267) zw48b03.r1"  
3719..3785  
misc\_feature /note="match to EST AU123510 (NID:g10948226)"  
3719..3785  
misc\_feature /note="match to EST BE272564 (NID:g9146913)"  
3719..3785  
misc\_feature /note="match to EST BE299708 (NID:g9183456)"  
3719..3785  
misc\_feature /note="match to EST BG766882 (NID:g14077535)"  
3719..3785  
misc\_feature /note="match to EST B1160365 (NID:g14620366)"  
3719..3785  
misc\_feature /note="similar to Mus musculus EST BE253526  
(NID:g8946272)"  
misc\_feature /note="match to EST AV686676 (NID:g10288539)"  
4730..4827  
repeat\_region /rpt\_family="L2"  
4834..4925  
repeat\_region /rpt\_family="L1"  
5715..5860  
misc\_feature /note="match to EST AA425526 (NID:g2106267) zw48b03.r1"  
5715..5860  
misc\_feature /note="match to EST AU123510 (NID:g10948226)"  
5715..5860  
misc\_feature /note="match to EST BE272564 (NID:g9146913)"  
5715..5860  
misc\_feature /note="match to EST BE299708 (NID:g9183456)"  
5715..5860  
misc\_feature /note="match to EST BG260659 (NID:g12770475)"  
5715..5860  
misc\_feature /note="match to EST BG766882 (NID:g14077535)"  
5715..5860  
misc\_feature /note="match to EST B1160365 (NID:g14620366)"  
5715..5860  
misc\_feature /note="similar to Mus musculus EST BE994936  
Query Match 100.0%; Score 96; DB 9; Length 149194;  
Best Local Similarity 100.0%; Pred. No. 3.4e-19;  
Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 GCCGCTGGCAGCTCTGAGGGAAGTATTAATATTAACACCATCTTACAGCTAGACCTCTT 60  
DB 90939 GCCGCTGGCAGCTCTGAGGGAAGTATTAATATTAACACCATCTTACAGCTAGACCTCTT 90880  
OY 61 TTGTAGAAAAGCAATGAGTGAAGTGCATTAAGT 96  
DB 90879 TTGTAGAAAAGCAATGAGTGAAGTGCATTAAGT 90844

RESULT 12  
AX000966 7582 bp DNA Linear PAT 10-MAR-2000  
LOCUS AX000966  
DEFINITION Sequence 11 from Patent WO9902696.  
ACCESSION AX000966  
VERSION AX000966.1 GI:7241208  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unclassified.  
REFERENCE 1 (bases 1 to 7582)  
AUTHORS Beseme,F. and Blond,J.  
TITLE ENDOGENETIC RETROVIRAL SEQUENCES, ASSOCIATED WITH AUTOIMMUNE  
DISEASES OR WITH PREGNANCY DISORDERS

JOURNAL Patent: WO 9902696-A 11 21-JAN-1999:  
BIO MERIEUX (FR); BESEME FREDERIC (FR)  
FEATURES  
source Location/Qualifiers  
1..7582  
/organism="unidentified"  
/db\_xref="taxon:32644"  
BASE COUNT 2156 a 1876 c 1538 g 1796 t 216 others  
ORIGIN  
Query Match 99.6%; Score 95.6; DB 6; Length 7582;  
Best Local Similarity 99.0%; Pred. No. 7.7e-19;  
Matches 95; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
OY 1 GCCGCTGGCAGCTCTGAGGGAAGTATTAATATTAACACCATCTTACAGCTAGACCTCTT 60  
DB 718 GCCGCTGGCAGCTCTGAGGGAAGTATTAATATTAACACCATCTTACAGCTAGACCTCTT 777  
OY 61 TTGTAGAAAAGCAATGAGTGAAGTGCATTAAGT 96  
DB 778 TTGTAGAAAAGCAATGAGTGAAGTGCATTAAGT 813

RESULT 13  
AX027480 7582 bp DNA Linear PAT 16-SEP-2000  
LOCUS AX027480  
DEFINITION Sequence 30 from Patent FR2788784.  
ACCESSION AX027480  
VERSION AX027480.1 GI:10188444  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 7582)  
AUTHORS Mallet,F., Voisset,C. and Paranhos,B.G.  
JOURNAL Patent: FR 2788784-A 30 28-JUL-2000;  
BIO MERIEUX (FR)  
FEATURES  
source Location/Qualifiers  
1..7582  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
BASE COUNT 2156 a 1876 c 1538 g 1796 t 216 others  
ORIGIN  
Query Match 99.6%; Score 95.6; DB 6; Length 7582;  
Best Local Similarity 99.0%; Pred. No. 7.7e-19;  
Matches 95; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
OY 1 GCCGCTGGCAGCTCTGAGGGAAGTATTAATATTAACACCATCTTACAGCTAGACCTCTT 60  
DB 718 GCCGCTGGCAGCTCTGAGGGAAGTATTAATATTAACACCATCTTACAGCTAGACCTCTT 777  
OY 61 TTGTAGAAAAGCAATGAGTGAAGTGCATTAAGT 96  
DB 778 TTGTAGAAAAGCAATGAGTGAAGTGCATTAAGT 813

RESULT 14  
AX000957 2938 bp DNA Linear PAT 10-MAR-2000  
LOCUS AX000957  
DEFINITION Sequence 2 from Patent WO9902696.  
ACCESSION AX000957  
VERSION AX000957.1 GI:7241199  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unclassified.  
REFERENCE 1 (bases 1 to 2938)  
AUTHORS Beseme,F. and Blond,J.  
TITLE ENDOGENETIC RETROVIRAL SEQUENCES, ASSOCIATED WITH AUTOIMMUNE  
DISEASES OR WITH PREGNANCY DISORDERS  
JOURNAL Patent: WO 9902696-A 2 21-JAN-1999;  
BIO MERIEUX (FR); BESEME FREDERIC (FR)

FEATURES Location/Qualifiers  
source 1..2938  
/organism="unidentified"  
/db\_xref="taxon:32644"

BASE COUNT 878 a 720 c 646 g 692 t 2 others

ORIGIN

Query Match 98.3%; Score 94.4; DB 6; Length 2938;  
Best Local Similarity 99.0%; Pred. No. 2.1e-18;  
Matches 95; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCCGCTGGCAGCTCTGAGGAGTATATATATACACCATCTAGACCTCTT 60  
|||||  
Db 133 GCCGCTGGCAGCTCTGAGGAGTATATATATACACCATCTAGACCTCTT 192  
|||||

QY 61 TTGTAGAAAAGCAATGAGTGAAGTGCCTAAGT 96  
|||||  
Db 193 TTGTAGAAAAGCAATGAGTGAAGTGCCTAAGT 228  
|||||

RESULT 15  
AX027471 2938 bp DNA linear PAT 16-SEP-2000  
LOCUS  
DEFINITION Sequence 21 from Patent FR2788784.  
ACCESSION AX027471  
VERSION AX027471.1 GI:10188435  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 2938)  
AUTHORS Mallet, F., Voisset, C. and Paranhos, B. G.  
JOURNAL Patent: FR 2788784-A 21 28-JUL-2000;  
BIO MERIEUX (FR)

FEATURES Location/Qualifiers  
source 1..2938  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"

BASE COUNT 878 a 720 c 646 g 692 t 2 others

ORIGIN

Query Match 98.3%; Score 94.4; DB 6; Length 2938;  
Best Local Similarity 99.0%; Pred. No. 2.1e-18;  
Matches 95; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCCGCTGGCAGCTCTGAGGAGTATATATATACACCATCTAGACCTCTT 60  
|||||  
Db 133 GCCGCTGGCAGCTCTGAGGAGTATATATATACACCATCTAGACCTCTT 192  
|||||

QY 61 TTGTAGAAAAGCAATGAGTGAAGTGCCTAAGT 96  
|||||  
Db 193 TTGTAGAAAAGCAATGAGTGAAGTGCCTAAGT 228  
|||||

Search completed: May 2, 2003, 12:53:45  
Job time : 467.145 secs



PT therapy of autoimmune disease, and abnormal or failed pregnancy  
XX  
PS Claim 1; Page 48-49; 106pp; French.  
XX  
CC This sequence represents clone cl.6n2 of the human endogenous retrovirus  
CC (HERV) W genome. The nucleic acids, their fragments or peptides encoded  
CC by them are markers of autoimmune disease (e.g. multiple sclerosis,  
CC rheumatoid polyarthritis, disseminated lupus erythematosus, insulin-  
CC dependent diabetes and related pathologies) and of abnormal or  
CC unsuccessful pregnancy and can be used as chromosomal markers for  
CC susceptibility to these conditions, or proximity markers of genes  
CC associated with this susceptibility.  
XX  
SQ Sequence 1321 BP; 365 A; 352 C; 277 G; 327 T; 0 other;  
Query Match 100.0%; Score 96; DB 20; Length 1321;  
Best Local Similarity 100.0%; Pred. No. 6.4e-24;  
Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GCCGCTGGCACTCTCTGAGGAGTATTAATTATACACCATCTTACAGCTAGACTCTT 60  
Db 717 GCCGCTGGCACTCTCTGAGGAGTATTAATTATACACCATCTTACAGCTAGACTCTT 776  
QY 61 TTGTAGAAAAGGCAATGAGTGAAGTCCATAAGT 96  
Db 777 TTGTAGAAAAGGCAATGAGTGAAGTCCATAAGT 812  
RESULT 2  
ID AAA59205 standard; DNA; 1321 BP.  
XX  
AC AAA59205;  
XX  
DT 07-NOV-2000 (first entry)  
XX  
DE 5' non coding and partial gag fragment of HERV-W from human genome.  
XX  
KW Autoimmune disease; retrovirus; human endogenous retrovirus W; HERV-W;  
XX gag gene; pregnancy; multiple sclerosis; T cell proliferation; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200043521-A2.  
XX  
PD 27-JUL-2000.  
XX  
PF 21-JAN-2000; 2000WO-FR00144.  
XX  
PR 21-JAN-1999; 99FR-0000888.  
XX  
PA (INMR) BIO MERIEUX.  
XX  
PI Paranhos-Baccala G, Mallet F, Voisset C;  
XX WPI; 2000-499229/44.  
XX  
DR WPI; 2000-499229/44.  
XX  
PT New nucleic acid from human endogenous retrovirus, useful e.g. for  
XX diagnosis of autoimmune disease and complications of pregnancy,  
XX PT contains at least part of the gag gene -  
XX  
XX  
PS Disclosure; Page 42-43; 53pp; French.  
XX  
CC The present sequence represents an endogenous retrovirus, useful e.g. for  
CC diagnosis of autoimmune disease, and is  
CC integrated into the human genome. The fragment is originally derived  
CC from a novel retrovirus, human endogenous retrovirus W (HERV-W). The  
CC HERV-W retrovirus is associated with autoimmune disease, failure of  
CC pregnancy or disorders of pregnancy. The nucleic acid fragment, or  
CC proteins derived from it, are useful for diagnosis of autoimmune  
CC disease (specifically multiple sclerosis) and for monitoring pregnancy.  
CC The nucleic acid fragments may also be used for in situ labelling of  
CC isolated chromosomes, while the transcription product can be used to

CC study or monitor T cell proliferation in vitro.  
XX  
SQ Sequence 1321 BP; 365 A; 352 C; 277 G; 327 T; 0 other;  
XX  
Query Match 100.0%; Score 96; DB 21; Length 1321;  
Best Local Similarity 100.0%; Pred. No. 6.4e-24;  
Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GCCGCTGGCACTCTCTGAGGAGTATTAATTATACACCATCTTACAGCTAGACTCTT 60  
Db 717 GCCGCTGGCACTCTCTGAGGAGTATTAATTATACACCATCTTACAGCTAGACTCTT 776  
QY 61 TTGTAGAAAAGGCAATGAGTGAAGTCCATAAGT 96  
Db 777 TTGTAGAAAAGGCAATGAGTGAAGTCCATAAGT 812  
RESULT 3  
ID ABN97928 standard; DNA; 1326 BP.  
XX  
AC ABN97928;  
XX  
DT 01-AUG-2002 (first entry)  
XX  
DE Human retroviral sequence gag.  
XX  
KW Autoimmune disease; HERV-7q; chromosome 7q; immunotherapy;  
XX multiple sclerosis; ds.  
XX  
OS Human retrovirus.  
XX  
PN WO967395-A1.  
XX  
PD 29-DEC-1999.  
XX  
PF 23-JUN-1999; 99WO-FR01513.  
XX  
PR 23-JUN-1998; 98FR-0007920.  
XX  
PA (INMR) INSERM INST NAT SANTE & RECH MEDICALE.  
XX  
PI Alliel PM, Perlin J, Rieger F;  
XX  
DR WPI; 2000-160587/14.  
XX  
PT New nucleic acid sequences of human endogenous retrovirus, HERV-7q,  
XX used for diagnosis, treatment and prevention of autoimmune and  
XX neurological diseases -  
XX  
XX  
PS Claim 2; Page 129; 225pp; French.  
XX  
CC The present invention relates to new nucleic acid sequences of human  
CC endogenous retrovirus, HERV-7q, which is located on chromosome 7q.  
CC Regulatory elements associated with HERV-7q may alter expression of other  
CC genes (even remote genes) on the same chromosome, inducing immunological  
CC and/or neurological changes (which may be pathological or protective/  
CC curative). HERV-7q peptides can be used to improve efficiency of the  
CC immune response, e.g. in immunotherapy. HERV-7q peptides and their coding  
CC sequences can be used in immunogenic or vaccinating compositions, for  
CC protection against autoimmune diseases, particularly multiple sclerosis.  
CC The peptides may also be used (by sequence comparison) to detect/identify  
CC endogenous retroviruses that are abnormally expressed in cancer,  
CC neuropathologies or other autoimmune diseases. The present sequence was  
CC used to illustrate the invention.  
XX  
SQ Sequence 1326 BP; 425 A; 306 C; 303 G; 292 T; 0 other;  
Query Match 100.0%; Score 96; DB 21; Length 1326;  
Best Local Similarity 100.0%; Pred. No. 6.4e-24;  
Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GCCGCTGGCACTCTCTGAGGAGTATTAATTATACACCATCTTACAGCTAGACTCTT 60

|||||  
Db 1 GCCGCTGGACCTCTGTGAGGAGATATAATTATTAACACCATCTTACAGCTAGACCTCTT 60  
OY 61 TTGTAGAAAAGGCAAAATGAGTGAAGTGCCTAATAGT 96  
|||||  
Db 61 TTGTAGAAAAGGCAAAATGAGTGAAGTGCCTAATAGT 96

RESULT 4  
AAS73701  
ID AAS73701 standard; cDNA: 1715 BP.  
XX  
AC AAS73701;  
XX  
DT 13-FEB-2002 (first entry)  
XX

DE DNA encoding novel human diagnostic protein #9505.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

OS Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

PE 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

PA (HYSE-) HXSEQ INC.

PI Drmanac RT, Liu C, Tang YT;

DR WPI; 2001-639362/73.

DR P-PSDB; ABG09514.

PT New isolated polynucleotide and encoded polypeptides, useful in  
diagnostics, forensics, gene mapping, identification of mutations  
responsible for genetic disorders or other traits and to assess  
biodiversity -

PS Claim 1; SEQ ID No 9505; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 1715 BP; 512 A; 435 C; 371 G; 397 T; 0 other;

Query Match 100.0%; Score 96; DB 23; Length 1715;  
Best Local Similarity 100.0%; Pred. No. 6.9e-24;  
Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCCGCTGGACCTCTGTGAGGAGATATAATTATTAACACCATCTTACAGCTAGACCTCTT 60  
|||||  
Db 655 GCCGCTGGACCTCTGTGAGGAGATATAATTATTAACACCATCTTACAGCTAGACCTCTT 714  
OY 61 TTGTAGAAAAGGCAAAATGAGTGAAGTGCCTAATAGT 96  
|||||  
Db 715 TTGTAGAAAAGGCAAAATGAGTGAAGTGCCTAATAGT 750

RESULT 5  
AAX25664  
ID AAX25664 standard; cDNA to mRNA: 2372 BP.  
XX  
AC AAX25664;  
XX  
DT 21-MAY-1999 (first entry)  
XX

DE Human endogenous retrovirus W clone cl.44.4.

KW Clone; human endogenous retrovirus; genome; autoimmune disease;  
KW multiple sclerosis; rheumatoid polyarthritis; insulin-dependent diabetes;  
KW disseminated lupus erythematosus; pregnancy; chromosomal marker; ss.

OS Human endogenous retrovirus.

PN WO9902696-A1.

PD 21-JAN-1999.

PF 06-JUL-1998; 98WO-FR01442.

PR 07-JUL-1997; 97FR-0008815.

PA (INMR) BIO MERIEUX.

PI Beseme F, Blond JL, Bouton O, Mallet F, Mandrand B;

DR WPI; 1999-120897/10.

PT New nucleic acid sequences from human endogenous retrovirus-W -  
expressed exclusively in placenta and useful in diagnosis and  
therapy of autoimmune disease, and abnormal or failed pregnancy

PS Claim 1; Page 68-70; 106pp; French.

XX This sequence represents clone cl.44.4 of the human endogenous retrovirus  
CC (HERV) W genome. The nucleic acids, their fragments or peptides encoded  
CC by them are markers of autoimmune disease (e.g. multiple sclerosis,  
CC rheumatoid polyarthritis, disseminated lupus erythematosus, insulin-  
CC dependent diabetes and related pathologies) and of abnormal or  
CC unsuccessful pregnancy and can be used as chromosomal markers for  
CC susceptibility to these conditions, or proximity markers of genes  
CC associated with this susceptibility.

XX Sequence 2372 BP; 674 A; 593 C; 544 G; 557 T; 4 other;

Query Match 100.0%; Score 96; DB 20; Length 2372;  
Best Local Similarity 100.0%; Pred. No. 7.3e-24;  
Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCCGCTGGACCTCTGTGAGGAGATATAATTATTAACACCATCTTACAGCTAGACCTCTT 60  
|||||  
Db 603 GCCGCTGGACCTCTGTGAGGAGATATAATTATTAACACCATCTTACAGCTAGACCTCTT 662

OY 61 TTGTAGAAAAGGCAAAATGAGTGAAGTGCCTAATAGT 96  
|||||  
Db 663 TTGTAGAAAAGGCAAAATGAGTGAAGTGCCTAATAGT 698

RESULT 6  
AAA59214  
ID AAA59214 standard; DNA: 2372 BP.

XX AC AAA59214;  
XX XX  
DT 07-NOV-2000 (first entry)  
XX XX  
DE R-U5 region and partial pol gene sequences of HERV-W from human genome.  
XX XX  
KM Autoimmune disease; retrovirus; human endogenous retrovirus W; HERV-W;  
XX gag gene; pregnancy; multiple sclerosis; T cell proliferation; ss.  
XX OS  
XX Homo sapiens.  
XX PN WO200043521-A2.  
XX XX  
PD 27-JUL-2000.  
XX XX  
PF 21-JAN-2000; 2000MO-FR00144.  
XX XX  
PR 21-JAN-1999; 99FR-0000888.  
XX XX  
PA (INMR ) BIO MERIEUX.  
XX XX  
PI Paranhos-Baccala G, Mallet F, Voisset C;  
XX DR WPI; 2000-499229/44.  
XX XX  
PT New nucleic acid from human endogenous retrovirus, useful e.g. for  
PT diagnosis of autoimmune disease and complications of pregnancy,  
PT contains at least part of the gag gene  
XX XX  
PS disclosure; Page 49; 53pp; French.  
XX XX  
CC The present sequence represents an endogenous retroviral nucleic acid  
CC fragment, which is associated with an autoimmune disease, and is  
CC integrated into the human genome. The fragment is originally derived  
CC from a novel retrovirus, human endogenous retrovirus W (HERV-W). The  
CC HERV-W retrovirus is associated with autoimmune disease, failure of  
CC pregnancy or disorders of pregnancy. The nucleic acid fragment, or  
CC proteins derived from it, are useful for diagnosis of autoimmune  
CC disease (specifically multiple sclerosis) and for monitoring pregnancy.  
CC The nucleic acid fragments may also be used for in situ labelling of  
CC isolated chromosomes, while the transcription product can be used to  
CC study or monitor T cell proliferation in vitro.  
XX XX  
SQ Sequence 2372 BP; 674 A; 593 C; 544 G; 557 T; 4 other;  
XX XX  
Query Match 100.0%; Score 96; DB 21; Length 2372;  
Best Local Similarity 100.0%; Pred. No. 7.5e-24;  
Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GCCGCGCTGGACATCTCGAGGAGATTAATTAACACCATCTTACAGCTAGACTCTT 60  
DB 603 GCCGCGCTGGACATCTCGAGGAGATTAATTAACACCATCTTACAGCTAGACTCTT 662  
QY 61 TTGTAGAAAAGCAATGAGTAGAGTGCCTAAGT 96  
DB 663 TTGTAGAAAAGCAATGAGTAGAGTGCCTAAGT 698  
XX XX  
RESULT 7  
AA576475  
ID AAS76475 standard; cDNA; 4349 BP.  
XX XX  
AC AAS76475;  
XX XX  
DT 13-FEB-2002 (first entry)  
XX XX  
DE DNA encoding novel human diagnostic protein #12279.  
XX XX  
KM Human: chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX OS  
XX Homo sapiens.

XX XX  
PN WO200175067-A2.  
XX XX  
PD 11-OCT-2001.  
XX XX  
PF 30-MAR-2001; 2001MO-US08631.  
XX XX  
PR 31-MAR-2000; 2000US-0540217.  
XX XX  
PR 23-AUG-2000; 2000US-0649167.  
XX XX  
PA (HYSE-) HYSEQ INC.  
XX XX  
PI Drmanac RT, Liu C, Tang YF;  
XX XX  
DR WPI; 2001-639362/73.  
XX XX  
DR P-PSDB; ABG12288.  
XX XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity  
XX XX  
PS Claim 1; SEQ ID No 12279; 103pp; English.  
XX XX  
CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
XX at ftp.wipo.int/pub/published\_pct\_sequences.  
XX XX  
SQ Sequence 4349 BP; 1319 A; 1061 C; 941 G; 1026 T; 2 other;  
XX XX  
Query Match 100.0%; Score 96; DB 23; Length 4349;  
Best Local Similarity 100.0%; Pred. No. 8.8e-24;  
Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GCCGCGCTGGACATCTCGAGGAGATTAATTAACACCATCTTACAGCTAGACTCTT 60  
DB 608 GCCGCGCTGGACATCTCGAGGAGATTAATTAACACCATCTTACAGCTAGACTCTT 667  
QY 61 TTGTAGAAAAGCAATGAGTAGAGTGCCTAAGT 96  
DB 668 TTGTAGAAAAGCAATGAGTAGAGTGCCTAAGT 703  
XX XX  
RESULT 8  
AAS68626  
ID AAS68626 standard; cDNA; 7466 BP.  
XX XX  
AC AAS68626;  
XX XX  
DT 13-FEB-2002 (first entry)  
XX XX  
DE DNA encoding novel human diagnostic protein #4430.  
XX XX  
KM Human: chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX OS  
XX Homo sapiens.



xx OS Homo sapiens.  
xx PN WO200175067-A2.  
xx PD 11-OCT-2001.  
xx PF 30-MAR-2001; 2001WO-US08631.  
xx PR 31-MAR-2000; 2000US-0540217.  
xx PR 23-AUG-2000; 2000US-0649167.  
xx PA (HYSE-) HYSEQ INC.  
xx PI Dmanac RT, Liu C, Tang YT;  
xx DR WPI: 2001-639362/73.  
xx DR P-PSDB: ABG04439.  
xx PT New isolated polynucleotide and encoded polypeptides, useful in  
xx PT diagnostics, forensics, gene mapping, identification of mutations  
xx PT responsible for genetic disorders or other traits and to assess  
xx PT biodiversity  
xx PS Claim 1; SEQ ID No 4430; 103pp; English.  
xx CC The invention relates to isolated polynucleotide (I) and  
xx CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
xx CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
xx CC and gene mapping, and in recombinant production of (II). The  
xx CC polynucleotides are also used in diagnostics as expressed sequence tags  
xx CC for identifying expressed genes. (I) is useful in gene therapy techniques  
xx CC to restore normal activity of (II) or to treat disease states involving  
xx CC (II). (II) is useful for generating antibodies against it, detecting or  
xx CC quantitating a polypeptide in tissue, as molecular weight markers and as  
xx CC a food supplement. (II) and its binding partners are useful in medical  
xx CC imaging of sites expressing (II). (I) and (II) are useful for treating  
xx CC disorders involving aberrant protein expression or biological activity.  
xx CC The polypeptide and polynucleotide sequences have applications in  
xx CC diagnostics, forensics, gene mapping, identification of mutations  
xx CC responsible for genetic disorders or other traits to assess biodiversity  
xx CC and to produce other types of data and products dependent on DNA and  
xx CC amino acid sequences. AAS64197-AAS94564 represent novel human  
xx CC diagnostic coding sequences of the invention.  
xx CC Note: The sequence data for this patent did not appear in the printed  
xx CC specification, but was obtained in electronic format directly from WIFO  
xx CC atftp.wifo.int/pub/published\_pcf\_sequences.  
xx SQ Sequence 7466 BP; 2217 A; 1880 C; 1614 G; 1754 T; 1 other;  
Query Match 100.0%; Score 96; DB 23; Length 7466;  
Best Local Similarity 100.0%; Pred. NO. 1e-23;  
Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GCCCGCTGGACCTCTGAGGAGATATAATTATACACATCTACAGCTAGACCTCT 60  
DB 4651 GCCCGCTGGACCTCTGAGGAGATATAATTATACACATCTACAGCTAGACCTCT 4710  
QY 61 TTGTAGAAAAGGCAATGAGTGAAGTGCCTAAGT 96  
DB 4711 TTGTAGAAAAGGCAATGAGTGAAGTGCCTAAGT 4746

RESULT 9  
ABN97929  
ID ABN97929 standard; DNA; 10499 BP.  
AC ABN97929;  
XX  
XX 01-AUG-2002 (first entry)  
DT  
XX Human retroviral sequence HERV-7q.  
DE  
XX

KW Autoimmune disease; HERV-7q; chromosome 7q; immunotherapy;  
xx KW multiple sclerosis; ds.  
xx OS Human retrovirus.  
xx PN WO9967395-A1.  
xx PD 29-DEC-1999.  
xx PF 23-JUN-1999; 99WO-FR01513.  
xx PR 23-JUN-1998; 98FR-0007920.  
xx PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.  
xx PI Alliel PM, Perin J, Rieger F;  
xx DR WPI: 2000-160587/14.  
xx PT New nucleic acid sequences of human endogenous retrovirus, HERV-7q,  
xx PT used for diagnosis, treatment and prevention of autoimmune and  
xx PT neurological diseases  
xx PS Claim 3; Fig 1; 225pp; French.  
xx CC The present invention relates to new nucleic acid sequences of human  
xx CC endogenous retrovirus, HERV-7q, which is located on chromosome 7q.  
xx CC Regulatory elements associated with HERV-7q may alter expression of other  
xx CC genes (even remote genes) on the same chromosome, inducing immunological  
xx CC and/or neurological changes (which may be pathological or protective/  
xx CC curative). HERV-7q peptides can be used to improve efficiency of the  
xx CC immune response, e.g. in immunotherapy. HERV-7q peptides and their coding  
xx CC sequences can be used in immunogenic or vaccinating compositions, for  
xx CC protection against autoimmune diseases, particularly multiple sclerosis.  
xx CC The peptides may also be used (by sequence comparison) to detect/identify  
xx CC endogenous retroviruses that are abnormally expressed in cancer,  
xx CC neuropathologies or other autoimmune diseases. The present sequence was  
xx CC used to illustrate the invention.  
xx SQ Sequence 10499 BP; 3048 A; 2676 C; 2280 G; 2495 T; 0 other;  
Query Match 100.0%; Score 96; DB 21; Length 10499;  
Best Local Similarity 100.0%; Pred. NO. 1.e-23;  
Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GCCCGCTGGACCTCTGAGGAGATATAATTATACACATCTACAGCTAGACCTCT 60  
DB 3065 GCCCGCTGGACCTCTGAGGAGATATAATTATACACATCTACAGCTAGACCTCT 3124  
QY 61 TTGTAGAAAAGGCAATGAGTGAAGTGCCTAAGT 96  
DB 3125 TTGTAGAAAAGGCAATGAGTGAAGTGCCTAAGT 3160

RESULT 10  
ABL61744  
ID ABL61744 standard; DNA; 56093 BP.  
AC ABL61744;  
XX  
XX 15-MAY-2002 (first entry)  
DT  
XX Colon adenocarcinoma related gene sequence SEQ ID NO:81.  
DE  
xx Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;  
xx stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;  
xx cytostatic; gene therapy; antineoplastic; Wilms tumour; adenocarcinoma;  
xx gene; ds.  
OS Homo sapiens.  
XX  
XX WO200194629-A2.  
PN  
XX

13-DEC-2001.  
30-MAY-2001; 2001WO-US10838.  
05-JUN-2000; 2000US-209473P.  
05-JUN-2000; 2000US-209531P.  
18-SEP-2000; 2000US-22133P.  
18-SEP-2000; 2000US-233617P.  
20-SEP-2000; 2000US-234009P.  
20-SEP-2000; 2000US-234034P.  
20-SEP-2000; 2000US-234052P.  
22-SEP-2000; 2000US-234509P.  
22-SEP-2000; 2000US-234567P.  
25-SEP-2000; 2000US-234923P.  
25-SEP-2000; 2000US-234924P.  
25-SEP-2000; 2000US-235077P.  
25-SEP-2000; 2000US-235082P.  
25-SEP-2000; 2000US-235134P.  
25-SEP-2000; 2000US-235280P.  
26-SEP-2000; 2000US-235637P.  
26-SEP-2000; 2000US-235638P.  
27-SEP-2000; 2000US-235711P.  
27-SEP-2000; 2000US-235720P.  
27-SEP-2000; 2000US-235840P.  
27-SEP-2000; 2000US-235863P.  
28-SEP-2000; 2000US-236028P.  
28-SEP-2000; 2000US-236032P.  
28-SEP-2000; 2000US-236033P.  
28-SEP-2000; 2000US-236034P.  
28-SEP-2000; 2000US-236109P.  
28-SEP-2000; 2000US-236111P.  
29-SEP-2000; 2000US-236842P.  
29-SEP-2000; 2000US-236891P.  
02-OCT-2000; 2000US-237172P.  
02-OCT-2000; 2000US-237173P.  
02-OCT-2000; 2000US-237278P.  
02-OCT-2000; 2000US-237294P.  
02-OCT-2000; 2000US-237295P.  
02-OCT-2000; 2000US-237316P.  
03-OCT-2000; 2000US-237425P.  
03-OCT-2000; 2000US-237598P.  
03-OCT-2000; 2000US-237604P.  
03-OCT-2000; 2000US-237606P.  
03-OCT-2000; 2000US-237608P.  
01-NOV-2000; 2000US-244867P.  
01-NOV-2000; 2000US-245084P.  
(AVAL- ) AVALON PHARM.  
Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;  
Soppet DR, Weaver Z;  
WPI; 2002-188264/24.  
Screening for anti-neoplastic agent involves exposing cells to a  
chemical agent to be tested for anti-neoplastic activity, and  
determining a change in expression of a gene of a signature gene set  
Claim 1; SEQ ID 81; 44pp; English.  
The present invention describes a method (M1) for screening for an  
anti-neoplastic agent. The method involves exposing cells to a chemical  
agent to be tested for anti-neoplastic activity, determining a change  
in expression of at least one gene (I) of a signature gene set, where (I)  
comprises a sequence (S) selected from 8447 sequences (given in ABL615666  
to ABL70110), or is at least 95% identical to (S), where a change in  
expression is indicative of anti-neoplastic activity. (I) has cytoskeletal  
activity and can be used in gene therapy. M1 can be used for screening  
an anti-neoplastic agent, and can be used for producing a product which  
is the data collected with respect to the anti-neoplastic agent as a  
result of M1, and the data is sufficient to convey the chemical  
structure and/or properties of the agent. M1 can be used in the  
treatment of cancer such as colon, breast, stomach, lung, thyroid,

Query Match	Best Local Similarity	Score 96:	DB 24:	Length 56093:
Matches 96: Conservative	100.0%;	100.0%;	Pred. No. 1,7e-23;	Mismatches 0; Indels 0; Gaps 0
1	GGCGCTGGCATTCCMGAGGAGATATATATATACACCACTTACCTGACCTCTT	60		
Db 31065	GGCGCTGGCATTCCMGAGGAGATATATATATATACACCACTTACCTGACCTCTT	31124		
61	TTGTAGAAAAGCAATGAGTGAAGTCCCAATAGT	96		
Db 31125	TTGTAGAAAAGCAATGAGTGAAGTCCCAATAGT	31160		
RESULT 11				
AAK25665				
ID	AAK25665 standard; cDNA to mRNA; 7582 BP.			
AC	AAK25665:			
XX	21-MAY-1999 (first entry)			
DE	Complete human endogenous retrovirus W genome.			
XX	Clone; human endogenous retrovirus; genome; autoimmune disease;			
KW	multiple sclerosis; rheumatoid polyarthritis; insulin-dependent diabetes;			
KW	disseminated lupus erythematosus; pregnancy; chromosomal marker; ss.			
XX	Human endogenous retrovirus.			
XX	WO9902696-A1.			
XX	21-JAN-1999.			
PD	06-JUL-1998; 98WO-FR01442.			
PF	07-JUL-1997; 97FR-0008815.			
PR	(INMR ) BIO MERIEUX.			
XX	Beseme F, Blond JL, Bouton O, Mallet F, Mandrand B;			
PI	WPI; 1999-120897/10.			
DR	New nucleic acid sequences from human endogenous retrovirus-W -			
PT	expressed exclusively in placenta and useful in diagnosis and			
PT	therapy of autoimmune disease, and abnormal or failed pregnancy			
XX	Claim 1; Page 71-74; 106pp; French.			
PS	This sequence represents the complete sequence of the human endogenous			
XX	retrovirus (HERV) W genome. The nucleic acids, their fragments or			
CC	peptides encoded by them are markers of autoimmune disease (e.g. multiple			
CC	sclerosis, rheumatoid polyarthritis, disseminated lupus erythematosus,			
CC	insulin-dependent diabetes and related pathologies) and of abnormal or			
CC	unsuccessful pregnancy and can be used as chromosomal markers for			
CC	susceptibility to these conditions, or proximity markers of genes			
CC	associated with this susceptibility.			
SO	Sequence 7582 BP; 2156 A; 1877 C; 1537 G; 1796 T; 2 U; 214 other:			
Query Match	99.6%;	Score 95.6;	DB 20;	Length 7582;
Best Local Similarity	99.0%;	Pred. No. 1.4e-23;		
Matches 95: Conservative	1; Mismatches 0; Indels 0; Gaps 0;			
1	GGCGCTGGCATTCCMGAGGAGATATATATATACACCACTTACCTGACCTCTT	60		

DB 718 GCCGCTGGCAGCTCCTGAGGAGATATTAATTAACACCATTTACAGCTAGACTCTT 777  
QY 61 TTGTAGAAAAGCAATGAGTGAAGTGCCTAAGT 96  
DB 778 TTGTAGAAAAGCAATGAGTGAAGTGCCTAAGT 813

RESULT 12  
AAA59215  
ID AAA59215 standard; DNA; 7582 BP.  
AC AAA59215;  
XX  
XX 07-NOV-2000 (first entry)  
DT  
XX Human endogenous retrovirus W (HERV-W) sequence.  
XX  
XX Autoimmune disease; retrovirus; human endogenous retrovirus W; HERV-W;  
KM gag gene; pregnancy; multiple sclerosis; T cell proliferation; ss.  
XX  
XX Human endogenous retrovirus.  
OS

Key Location/Qualifiers  
FH 1..120  
FT LTR  
FT /\*tag= a  
FT /note= "R of 5' LTR"  
FT 121..575  
FT LTR  
FT /\*tag= b  
FT /note= "U5 of 5' LTR"  
FT 579..596  
FT primer\_bind  
FT /\*tag= c  
FT /note= "ORF1 env538"  
FT 5581..7194  
FT CDS  
FT /\*tag= d  
FT /note= "ORF1 env538"  
FT 7039..7194  
FT CDS  
FT /\*tag= e  
FT /note= "ORF2 52 AA"  
FT 7112..7255  
FT CDS  
FT /\*tag= f  
FT /note= "ORF3 48 AA"  
FT 7244..7254  
FT misc-feature  
FT /\*tag= g  
FT /note= "polyurine tract"  
FT 7256..7582  
FT LTR  
FT /\*tag= h  
FT /note= "U3-R of 3' LTR"  
FT 7563..7569  
FT polyA\_signal  
FT /\*tag= i  
FT  
PN WO200043521-A2.  
XX  
XX 27-JUL-2000.  
PD  
XX  
XX 21-JAN-2000; 2000WO-FR00144.  
PF  
XX 21-JAN-1999; 99FR-0000888.  
PR  
XX  
XX (INMR ) BIO MERIEUX.  
PA  
XX Paranhos-Baccala G, Mallet F, Voliset C;  
PI  
XX WPI; 2000-499229/44.  
DR  
XX  
XX New nucleic acid from human endogenous retrovirus, useful e.g. for  
PT diagnosis of autoimmune disease and complications of pregnancy,  
PT contains at least part of the gag gene  
XX  
XX  
PS Disclosure; Page 49-52; 53pp; French.  
XX  
XX The present sequence represents an endogenous retrovirus, which is  
CC associated with an autoimmune disease, and is integrated into the human  
CC genome. The retrovirus is human endogenous retrovirus W (HERV-W). The  
CC HERV-W retrovirus is associated with autoimmune disease, failure of

CC pregnancy or disorders of pregnancy. HERV-W nucleic acid fragments, or  
CC proteins derived from it, are useful for diagnosis of autoimmune  
CC disease (specifically multiple sclerosis) and for monitoring pregnancy.  
CC The nucleic acid fragments may also be used for in situ labelling of  
CC isolated chromosomes, while the transcription product can be used to  
CC study or monitor T cell proliferation in vitro.  
XX

Sequence 7582 BP; 2156 A; 1876 C; 1538 G; 1796 T; 216 other;  
SQ

Query Match 99.6%; Score 95.6; DB 21; Length 7582;  
Best Local Similarity 99.0%; Pred. No. 1.4e-23;  
Matches 95; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 GCCGCTGGCAGCTCCTGAGGAGATATTAATTAACACCATTTACAGCTAGACTCTT 60  
DB 718 GCCGCTGGCAGCTCCTGAGGAGATATTAATTAACACCATTTACAGCTAGACTCTT 777

QY 61 TTGTAGAAAAGCAATGAGTGAAGTGCCTAAGT 96  
DB 778 TTGTAGAAAAGCAATGAGTGAAGTGCCTAAGT 813

## RESULT 13

AAX25656

ID AAX25656 standard; CDNA to mRNA; 2938 BP.

XX AAX25656;

XX 21-MAY-1999 (first entry)

XX Human endogenous retrovirus W clone cl.6A1.

XX Clone; human endogenous retrovirus; genome; autoimmune disease;

XX multiple sclerosis; rheumatoid polyarthritis; insulin-dependent diabetes;

XX disseminated lupus erythematosus; pregnancy; chromosomal marker; ss.

XX Human endogenous retrovirus.

XX WO9902696-A1.

XX 21-JAN-1999.

XX 06-JUL-1998; 98WO-FR01442.

XX 07-JUL-1997; 97FR-0008815.

XX (INMR ) BIO MERIEUX.

XX Beseme F, Blond JL, Bouton O, Mallet F, Mandrand B;

XX WPI; 1999-120897/10.

XX New nucleic acid sequences from human endogenous retrovirus-W -

XX expressed exclusively in placenta and useful in diagnosis and

XX therapy of autoimmune disease, and abnormal or failed pregnancy

XX Claim 1; Page 49-52; 106pp; French.

XX This sequence represents clone cl.6A1 of the human endogenous retrovirus

XX (HERV) W genome. The nucleic acids, their fragments or peptides encoded

XX by them are markers of autoimmune disease (e.g. multiple sclerosis,

XX rheumatoid polyarthritis, disseminated lupus erythematosus, insulin-

XX dependent diabetes and related pathologies) and of abnormal or

XX unsuccessful pregnancy and can be used as chromosomal markers for

XX susceptibility to these conditions, or proximity markers of genes

XX associated with this susceptibility.

SQ Sequence 2938 BP; 878 A; 720 C; 646 G; 692 T; 2 other;

Query Match 98.3%; Score 94.4; DB 20; Length 2938;

Best Local Similarity 99.0%; Pred. No. 2.9e-23;

Matches 95; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCCGCTGGCACTCTGTAGGAGATATAATTAACACCATCTTACAGCTAGACCTCTT 60  
DB 133 GCCGCTGGCACTCTGTAGGAGATATAATTAACACCATCTTACAGCTAGACCTCTT 192  
QY 61 TTGTAGAAAAGCAATGAGTGAAGTCCATTAAGT 96  
DB 193 TTGTAGAAAAGCAATGAGTGAAGTCCATTAAGT 228

## RESULT 14

AAA59206  
ID AAA59206 standard; DNA; 2938 BP.

AAA59206;

07-NOV-2000 (first entry)

XX Gag and partial pol gene fragment of HENV-W from human genome.

KW Autoimmune disease; retrovirus; human endogenous retrovirus W; HENV-W;  
XX gag gene; pregnancy; multiple sclerosis; T cell proliferation; ss.

OS Homo sapiens.

XX MO200043521-A2.

XX 27-JUL-2000.

XX 21-JAN-2000; 2000WO-FR00144.

XX 21-JAN-1999; 99FR-0000888.

XX (1NMR) BIO MERIEUX.

XX Paranhos-Baccala G, Mallet F, Voliset C;

XX WPI: 2000-499229/44.

PT New nucleic acid from human endogenous retrovirus, useful e.g. for  
PT diagnosis of autoimmune disease and complications of pregnancy,  
PT contains at least part of the gag gene

XX Disclosure; Page 43; 53pp; French.

XX The present sequence represents an endogenous retroviral nucleic acid  
CC fragment, which is associated with an autoimmune disease, and is  
CC integrated into the human genome. The fragment is originally derived  
CC from a novel retrovirus, human endogenous retrovirus W (HENV-W). The  
CC HENV-W retrovirus is associated with autoimmune disease, failure of  
CC pregnancy or disorders of pregnancy. The nucleic acid fragment, or  
CC proteins derived from it, are useful for diagnosis of autoimmune  
CC disease (specifically multiple sclerosis) and for monitoring pregnancy.  
CC The nucleic acid fragments may also be used for in situ labelling of  
CC isolated chromosomes, while the transcription product can be used to  
CC study or monitor T cell proliferation in vitro.

XX Sequence 2938 BP; 878 A; 720 C; 646 G; 692 T; 2 other;

Query Match 98.3%; Score 94.4; DB 21; Length 2938;

Best Local Similarity 99.0%; Pred. No. 2.9e-23;

Matches 95; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GCCGCTGGCACTCTGTAGGAGATATAATTAACACCATCTTACAGCTAGACCTCTT 60

DB 133 GCCGCTGGCACTCTGTAGGAGATATAATTAACACCATCTTACAGCTAGACCTCTT 192

QY 61 TTGTAGAAAAGCAATGAGTGAAGTCCATTAAGT 96

DB 193 TTGTAGAAAAGCAATGAGTGAAGTCCATTAAGT 228

## RESULT 15

AA584210

ID AA584210 standard; CDNA; 6394 BP.

XX AA584210;

XX 13-FEB-2002 (first entry)

XX DNA encoding novel human diagnostic protein #20014.

KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX Homo sapiens.

XX MO200175067-A2.

XX 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US08631.

XX 31-MAR-2000; 2000US-0540217.

XX 23-AUG-2000; 2000US-0649167.

XX (HXSE-) HXSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

XX WPI: 2001-639362/73.

XX P-PSDB: ABG20023.

PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity

XX Claim 1; SEQ ID NO 20014; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and  
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,  
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
XX and gene mapping, and in recombinant production of (II). The  
XX polynucleotides are also used in diagnostics as expressed sequence tags  
XX for identifying expressed genes. (I) is useful in gene therapy techniques  
XX to restore normal activity of (II) or to treat disease states involving  
XX (II). (II) is useful for generating antibodies against it, detecting or  
XX quantitating a polypeptide in tissue, as molecular weight markers and as  
XX a food supplement. (II) and its binding partners are useful in medical  
XX imaging of sites expressing (II). (I) and (II) are useful for treating  
XX disorders involving aberrant protein expression or biological activity.  
XX The polypeptide and polynucleotide sequences have applications in  
XX diagnostics, forensics, gene mapping, identification of mutations  
XX responsible for genetic disorders or other traits to assess biodiversity  
XX and to produce other types of data and products dependent on DNA and  
XX amino acid sequences. AA584197-AA594564 represent novel human  
XX diagnostic coding sequences of the invention.  
XX Note: The sequence data for this patent did not appear in the printed  
XX specification, but was obtained in electronic format directly from WIPO  
XX at ftp.wipo.int/pub/published\_pcl\_sequences.

XX Sequence 6394 BP; 1840 A; 1597 C; 1384 G; 1571 T; 2 other;

Query Match 79.8%; Score 76.6; DB 23; Length 6394;

Best Local Similarity 92.9%; Pred. No. 6.6e-17;

Matches 92; Conservative 0; Mismatches 4; Indels 3; Gaps 1;

QY 1 GCCGCTGGCACTCTGTAGGAGATATAATTAACACCATCTTACAGCTAGACCTCTT 60

DB 2697 GCCGCTGGCACTCTGTAGGAGATATAATTAACACCATCTTACAGCTAGACCTCTT 2756

QY 61 TTGTAGAAAAGCAATGAGTGAAGTCCATTAAGT 96

DB 2757 TTGTAGAAAAGCAATGAGTGAAGTCCATTAAGT 2795

Mon May 5 08:41:35 2003

us-09-719-554-3\_copy\_3065\_3160.rng

Page 9

Search completed: May 2, 2003, 11:29:57  
Job time : 172.938 secs

---



GenCore version 5.1.5  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 2, 2003, 11:21:20 ; Search time 1105.66 seconds

(without alignments)  
1406.193 Million cell updates/sec

Title: US-09-719-554-3\_COPY\_3065\_3160

Perfect score: 96  
Sequence: 1 gccgcctgcgacctcctgagg.....tggagtgagtgccataagt 96

Scoring table: IDENTITY\_NUC  
Gapop:10.0, Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estnu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_estc:\*  
9: gb\_estl1:\*  
10: gb\_estl2:\*  
11: gb\_est3:\*  
12: gb\_est4:\*  
13: gb\_est5:\*  
14: gb\_est6:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vit:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_fod:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	96	100.0	450	B0024410
2	68.8	71.7	913	B0424090
3	67.2	70.0	681	AG009944
4	67.2	70.0	685	AG009932
5	67.2	70.0	689	AG009943
6	67.2	70.0	692	AG009942

7	66.6	69.4	711	AG009923
8	64.6	67.3	401	AG014538
9	62.4	65.0	690	BM980860
10	61.4	64.0	559	A0833457
11	49.6	51.7	400	HS_5296_B
12	48	50.0	354	A0058957
13	48	50.0	373	A0110624
14	48	50.0	851	A0748718
15	47.6	49.6	632	A0000365
16	46.4	48.3	429	A0035279
17	46.4	48.3	559	AL708722
18	46	47.9	237	B45372
19	44.8	46.7	545	A0133205
20	44.8	46.7	711	AV730914
21	43.4	45.2	420	A0435686
22	43.4	45.2	473	A0248325
23	43.2	45.0	444	B99733
24	43.2	45.0	493	B91430
25	42.8	44.6	358	B03917
26	42.6	44.4	513	A0034931
27	42.6	44.4	645	AL718617
28	42.6	44.4	500	B53643
29	42.2	44.0	500	B53275
30	41.8	43.5	318	AA904177
31	41.6	43.3	314	AA983698
32	41.6	43.3	358	AT121949
33	41.6	43.3	370	AM105235
34	41.6	43.3	378	BF960517
35	41.6	43.3	408	B98887
36	41.6	43.3	454	B60060
37	41.6	43.3	465	BM718032
38	41.6	43.3	465	BM680485
39	41.6	43.3	543	B71000
40	41.6	43.3	568	B55061
41	41.6	43.3	602	B55062
42	41.6	43.3	632	A0021694
43	41.6	43.3	584	A0351335
44	41	42.7	477	A0224121
45	40.6	42.3	17	A0224121

#### ALIGNMENTS

RESULT 1  
B0024410/c 450 bp mRNA linear EST 27-MAR-2002  
LOCUS  
DEFINITION  
UI-1-BBIP-aug-e-06-0-UI.s1 NCI-CCAP P16 Homo sapiens CDNA clone  
UI-1-BBIP-aug-e-06-0-UI 3', mRNA sequence.  
ACCESSION  
B0024410  
VERSION  
B0024410.1 GI:19759689  
KEYWORDS  
EST.  
SOURCE  
human.  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
1 (bases 1 to 450)  
AUTHORS  
NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.  
TITLE  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
JOURNAL  
Unpublished (1997)  
COMMENT  
Contact: Robert Strausberg, Ph.D.  
Email: cga@bbs-remail.nih.gov  
Tissue Procurement: Dr. Steven Brown  
CDNA Library preparation: Dr. M. Bento Soares, University of Iowa  
CDNA library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Clone distribution information can be obtained  
from Dr. M. Bento Soares, bento-soares@uiowa.edu  
The following repetitive elements were found in this CDNA  
sequence: 14-226, >HERV17#LTR/Retroviral (matched complement)  
214-256, >HERV17#LTR/Retroviral (matched complement) 257-450,  
>LTR17#LTR/Retroviral (matched complement)

Seq primer: M13 FORWARD  
FEATURES  
POLY-A-Yes  
Location/Qualifiers  
1. 450

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="UI-1-BB1P-aug-e-06-0-UI"  
/clone\_lib="NCI\_CGAP\_P16"  
/tissue\_type="Placenta"  
/dev\_stage="Full Term"  
/lab\_host="DH10B (Life Technologies)"  
/note="Organ: Placenta; Vector: pRT73-Pac (Pharmacia) with a modified polylinker; Site\_1: EcoR I; Site\_2: Not I; NCI\_CGAP\_P16 is a subtracted cDNA library constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pRT73-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The additional information, contact: Bento Soares, bento-soares@uiowa.edu  
TAG\_LIB=UI-1-BB1P  
TAG\_TISSUE=Placenta human full term  
TAG\_SEQ=AGGAA"

BASE COUNT 94 a 96 c 113 g 147 t  
ORIGIN

Query Match 100.0%; Score 96; DB 14; Length 450;  
Best Local Similarity 100.0%; Pred. 1.2e-20;  
Matches 96; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GCCGCTGGACCTCTGAGGAGTAATTAATTAACACCATCTTACAGTACCTCTT 60  
Db 114 GCCGCTGGACCTCTGAGGAGTAATTAATTAACACCATCTTACAGTACCTCTT 55

OY 61 TTGTAGAAAAGCAATGAGTGAAGTGCCTAAGT 96  
Db 54 TTGTAGAAAAGCAATGAGTGAAGTGCCTAAGT 19

RESULT 2  
BQ424090 913 bp mRNA linear EST 23-MAY-2002  
LOCUS BQ424090  
DEFINITION AGENCOURT 7892347 NIH\_MGC\_72 Homo sapiens cDNA clone IMAGE:6158499  
5', mRNA sequence.  
ACCESSION BQ424090.1 GI:21119405  
VERSION BQ424090.1  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 913)  
NIH-MGC http://mgi.nci.nih.gov/  
National Institutes of Health, Mammalian Gene Collection (MGC)  
Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: ATCC/DC/DTP  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: L1AM13506 row: b column: 04  
High quality sequence stop: 619.  
Location/Qualifiers  
1. 913

FEATURES  
source

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6158499"  
/clone\_lib="NIH\_MGC\_72"  
/tissue\_type="melanotic melanoma"  
/lab\_host="DH10B (Phage-resistant)"  
/note="Organ: skin; Vector: pCMV-SPORT6; Site\_1: NotI; Site\_2: SalI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 2 kb. Library constructed by Life Technologies."  
BASE COUNT 238 a 267 c 196 g 212 t  
ORIGIN

Query Match 71.7%; Score 68.8; DB 14; Length 913;  
Best Local Similarity 94.3%; Pred. No. 9.2e-12;  
Matches 83; Conservative 0; Mismatches 2; Indels 3; Gaps 1;

OY 12 CTCCTGAGGAGTAATTAATTAACACCATCTTACAGTACCTCTTTGTAGAA--- 68  
Db 634 CTCCTGAGGAGTAATTAATTAACACCATCTTACAGTACCTCTTTGTAGAAAG 693

OY 69 AAGCAATGAGTGAAGTGCCTAAGT 96  
Db 694 AAGCAATGAGTGAAGTGCCTAAGT 721

RESULT 3  
AG009944/c 681 bp DNA linear GSS 14-APR-1999  
LOCUS AG009944/c  
DEFINITION Homo sapiens genomic DNA, 21q region, clone: 72018A32, genomic survey sequence.  
ACCESSION AG009944 AG003755  
VERSION AG009944.1 GI:3289930  
KEYWORDS GSS.  
SOURCE Homo sapiens DNA, clone: 72018A32.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 681)  
Hattori, M., Ishii, K., Toyoda, A., Shiba, T. and Sakaki, Y.  
Homo sapiens genomic DNA, chromosome 21q  
Published only in Database (1998)  
REFERENCE  
Hattori, M., Ishii, K., Toyoda, A., Shiba, T. and Sakaki, Y.  
Direct Submission  
Submitted (06-JUL-1998) Masahira Hattori, Kitasato University,  
Department of Science, JST Sequencing Laboratory, Kitasato 1-15-1,  
Sagamihara 228, Japan (E-mail: hattori@hgc.ims.u-tokyo.ac.jp,  
Tel: 0427-78-9732, Fax: 0427-78-9561)  
On Feb 5, 1999 this sequence version replaced g1:2754657.  
AG003755: Submitted (07-Jan-1998).  
COMMENT  
Location/Qualifiers  
1. 681  
source  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="21"  
/map="21q"  
/clone="72018A32"

BASE COUNT 141 a 143 c 180 g 194 t 23 others  
ORIGIN

Query Match 70.0%; Score 67.2; DB 17; Length 681;  
Best Local Similarity 93.2%; Pred. No. 2.8e-11;  
Matches 82; Conservative 0; Mismatches 3; Indels 3; Gaps 1;

OY 12 CTCCTGAGGAGTAATTAATTAACACCATCTTACAGTACCTCTTTGTAGAA---A 68  
Db 442 CCCCTGAGGAGTAATTAATTAACACCATCTTACAGTACCTCTTTGTAGAAAG 383

OY 69 AAGCAATGAGTGAAGTGCCTAAGT 96  
Db 382 AAGCAATGAGTGAAGTGCCTAAGT 355



```

RESULT 4
AG009932/c AG009932      685 bp       DNA           linear    GSS 14-APR-1999
LOCUS              Homo sapiens genomic DNA, 21q region, clone: 72018A32, genomic
DEFINITION          survey sequence.
ACCESSION            AG009932
VERSION              AG009932.1 GI:3289918
KEYWORDS             GSS.
SOURCE               Homo sapiens DNA, clone:72018A32.
ORGANISM             Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE            1 (bases 1 to 685)
AUTHORS             Hattori,M., Ishii,K., Toyoda,A., Shiba,T. and Sakaki,Y.
JOURNAL              Homo sapiens genomic DNA, chromosome 21q
TITLE                Published Only in Database (1998)
REFERENCE            2 (bases 1 to 685)
AUTHORS             Hattori,M., Ishii,K., Toyoda,A., Shiba,T. and Sakaki,Y.
JOURNAL              Direct Submission
TITLE                Submitted (06-JUL-1998) Masahira Hattori, Kitasato University,
                    Department of Science, JST Sequencing Laboratory; Kitasato I-15-1,
                    Sagami-hara 228, Japan (E-mail:hattorileg@cc.kit.ac.jp),
                    Tel.:0427-78-9732, Fax:0427-78-9561)
COMMENT              On Feb 5, 1999 this sequence version replaced gi:2754645.
FEATURES
Source              Location/Oualifiers
                   1..685
                   /organism="Homo sapiens"
                   /db_xref="taxon:9606"
                   /chromosome="21"
                   /map="21q"
BASE COUNT          148 a     144 c     190 g     199 t     4 others
ORIGIN
Query Match        70.0%; Score 67.2; DB 17; Length 685;
Best Local Similarity 93.2%; Pred.No.2.8e-11;
Matches 82; Conservative 0; Mismatches 3; Indels 3; Gaps 1;
QY   12 CTCCTGAGGGAGAATAATTAATATATAACACCATTTCACACTAGACCTCTTTGTAGA--A 68
Db   440 CCCCTGAGCGGAAGTAAATATATATAACACCACTCCTACACTGCACCTCTTTGTAGAAA GA 381
QY   69 AAGGCAATGAGAGTGCAATGCCCATAGT 96
Db   380 AGGGCAATGAGAGTGCAATGCCCATATGT 353

RESULT 5
AG009943/c AG009943      689 bp       DNA           linear    GSS 14-APR-1999
LOCUS              Homo sapiens genomic DNA, 21q region, clone: 72018A32, genomic
DEFINITION          survey sequence.
ACCESSION            AG009943
VERSION              AG003754
KEYWORDS             AG009943.1 GI:3289929
SOURCE               GSS.
ORGANISM             Homo sapiens DNA, clone:72018A32.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE            1 (bases 1 to 689)
AUTHORS             Hattori,M., Ishii,K., Toyoda,A., Shiba,T. and Sakaki,Y.
JOURNAL              Homo sapiens genomic DNA, chromosome 21q
TITLE                Published Only in Database (1998)
REFERENCE            2 (bases 1 to 689)
AUTHORS             Hattori,M., Ishii,K., Toyoda,A., Shiba,T. and Sakaki,Y.
JOURNAL              Direct Submission
TITLE                Submitted (06-JUL-1998) Masahira Hattori, Kitasato University,
                    Department of Science, JST sequencing Laboratory; Kitasato I-15-1,
                    Sagami-hara 228, Japan (E-mail:hattorilegc@cc.kit.ac.jp),
                    Tel.:0427-78-9732, Fax:0427-78-9561)
COMMENT              On Feb 5, 1999 this sequence version replaced gi:2754656.

```

[illegible]

Email: jwallace@u.washington.edu  
Clones are derived from the human BAC library RPCT-11. For BAC library availability, please contact Pieter de Jong (pietere@jorg.med.buffalo.edu). Clones may be purchased from BACPAC Resources ([http://Bacpac.med.buffalo.edu/ordering\\_bac.htm](http://Bacpac.med.buffalo.edu/ordering_bac.htm)) or from Resear h Genetics (info@resgen.com). BAC end Web Server: <http://www.hsc.washington.edu>  
Plate: 701 row: B column: 6  
Seq primer: SP6  
Class: BAC ends  
High quality sequence stop: 401.

FEATURES  
source  
Location/Qualifiers  
1..401  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_1fb="PACI-11 Human Male BAC Library"  
/sex="male"  
/note="Vector: pBAC3.6; Site\_1: EcoRI; Site\_2: EcoRI; Male blood DNA was isolated from one randomly chosen donor and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBAC3.6 vector at EcoRI sites"

BASE COUNT  
95 a 77 c 87 g 140 t 2 others

ORIGIN  
Query Match 67.3%; Score 64.6; DB 17; Length 401;  
Best Local Similarity 90.9%; Pred. No. 1.6e-10;  
Matches 80; Conservative 0; Mismatches 5; Indels 3; Gaps 1;

OY 12 CTCCTGAGGAGAACTTAATATAACACCATTCTTACACTGACCTCTTTGTAGA---A 68  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 325 CCCTNGAGGAAGTAATTAATTAACACCACTTACAGCCTTGACCTCTTTGTAGAAAGA 266

OY 69 AAGCGAATGGAGTGAACTGCCATTAAGT 96  
| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
Db 265 AGGCGAATGGAGTGAACTGCCATTAAGT 238

RESULT 9  
BM980860/c 690 bp - mRNA linear EST 21-MAR-2002  
LOCUS BM980860  
DEFINITION UI-CF-EN1-a-de-g-24-0-UI\_51 UI-CF-EN1 Homo sapiens cDNA clone  
UI-CF-EN1-a-de-g-24-0-UI\_3, mRNA sequence.  
ACCESSION BM980860  
VERSION BM980860.1 GI:19602750  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE 1 (bases 1 to 690)  
AUTHORS Ronaldo,M.F., Lennon,G. and Soares,M.B.  
TITLE Normalization and subtraction: two approaches to facilitate gene discovery  
Genome Res. 6 (9), 791-806 (1996)  
JOURNAL MEDLINE 97044477  
COMMENT Contact: McCray, PB  
McCray Lab  
University of Iowa  
2024 University of Iowa Med Labs, Iowa City, IA 52242, USA  
Tel: 319 356 4866  
Fax: 319 356 7171  
Email: paul-mccray@uiowa.edu  
Tissue Procurement: Dr. M. J. Welsh, University of Iowa  
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa  
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Researchers may obtain clones from Research Genetics ([www.resgen.com](http://www.resgen.com)).  
The following repetitive elements were found in this cDNA sequence: 10-148, >HERV1/LTR/Retroviral (matched complement) 18-266, >HEBV9/LTR/Retroviral (matched complement) 176-296.

>HERV17#LTR/Retroviral (matched compliment) 327-690,  
>LTR17#LTR/Retroviral (matched compliment)

Seq primer: M13 FORWARD  
POLY-A-yes.

# FEATURES

## SOURCE

Location/Qualifiers  
1. 690  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="UT-CF-EN1-ade-g-24-0-UT"  
/clone\_lib="UT-CF-EN1"  
/tissue\_type="Primary Lung Cystic Fibrosis Epithelial Cells"  
/dev\_stage="Adult"  
/lab\_host="DH10B (Life Technologies) (T1 phage resistant)"  
/note="Organ: Lung; Vector: pT73-Pac (Pharmacia) with a modified polylinker; Site\_1: EcoR I; Site\_2: Not I; UT-CF-EN1 is a normalized cDNA library containing the following tissue(s): Primary Lung Cystic Fibrosis Epithelial Cells. The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT73-Pac vector. The oligonucleotide used to prime the synthesis of the first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is CTGCTCAGGT.  
TAG\_LIB=UT-CF-EN1  
TAG\_TISSUE=Human Lung Epithelial cell lines untreated LPS 6hr to LPS 24h  
TAG\_SEQ=CTGCTCAGGT"

BASE COUNT 151 a 155 c 191 g 193 t

## ORIGIN

Query Match 65.0%; Score 62.4; DB 14; Length 690;  
Best Local Similarity 89.8%; Pred. No. 1e-09;  
Matches 79; Conservative 0; Mismatches 6; Indels 3; Gaps 1;

QY 12 CTCTGAGGGAATTAATTAATTAACACCACTTACAGCTACCTCTTTGTAGAA--- 68  
|||||  
DB 142 CTCTGAGGGAATTAATTAATTAACACTTACAGCTACCTCTTTGTAGAAAG 83  
QY 69 AAGCAATGAGTGAAGTGCATATAGT 96  
|||||  
DB 82 AAGCAATGAGTGAAGTGCATATAGT 55

## RESULT 10

LOCUS A0833457

DEFINITION HS\_5596\_B2\_H12\_T7A RPO1-11 Human Male BAC Library Homo sapiens  
559 bp DNA linear GSS 27-AUG-1999  
A0833457  
genomic clone Plate=872 Col=24 Row=P, DNA sequence.

ACCESSION A0833457  
VERSION A0833457.1 GI:5799442  
KEYWORDS GSS.

SOURCE human.  
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 559)  
AUTHORS Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Kellier,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D., and Hood,L.

TITLE Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)  
MEDLINE 99380589

COMMENT Contact: Mahairas GG, Wallace JC, Hood L  
High Throughput Sequencing Center  
University of Washington  
401 Queen Anne Avenue North, Seattle, WA 98109, USA  
Tel: (206) 616-3618

Fax: (206) 616-3887

Email: jwallace@u.washington.edu

Clones are derived from the human BAC library RPO1-11. For BAC library availability, please contact Pieter de Jong (pieterdejong.med.bufileo.edu). Clones may be purchased from BACPAC Resources ([http://bacpac.med.bufileo.edu/ordering\\_bac.htm](http://bacpac.med.bufileo.edu/ordering_bac.htm)) or from Research Genetics (info@resgen.com). BAC end Web Server: <http://www.htsc.washington.edu>  
Plate: 872 Row: P Column: 24

Seq primer: T7

Class: BAC ends

High quality sequence stop: 559.

# FEATURES

## SOURCE

Location/Qualifiers  
1. 559  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="Plate=872 Col=24 Row=P"  
/clone\_lib="RPO1-11 Human Male BAC Library"  
/sex="male"  
/note="Vector: pBAC3.6; Site\_1: EcoRI; Site\_2: EcoRI; Male blood DNA was isolated from one randomly chosen donor and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBAC3.6 vector at EcoRI sites"

BASE COUNT 161 a 146 c 117 g 122 t 13 others

## ORIGIN

Query Match 64.0%; Score 61.4; DB 17; Length 559;  
Best Local Similarity 88.6%; Pred. No. 2e-09;  
Matches 78; Conservative 0; Mismatches 7; Indels 3; Gaps 1;

QY 12 CTCTGAGGGAATTAATTAATTAACACCACTTACAGCTACCTCTTTGTAG---AA 68  
|||||  
DB 328 CTCTGAGGGAATTAATTAATTAACACCACTTACAGCTACCTCTTTGTAGAGAG 387  
QY 69 AAGCAATGAGTGAAGTGCATATAGT 96  
|||||  
DB 388 AAGCAATGAGTGAAGTGCATATAGT 415

## RESULT 11

LOCUS T59288

DEFINITION T59288 400 bp mRNA linear EST 09-FEB-1995  
IMAGE:79506 5', mRNA sequence.

ACCESSION T59288  
VERSION T59288.1 GI:661125  
KEYWORDS EST.

SOURCE human.  
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 400)

AUTHORS Haller,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B., Chissos,S., Dietrich,N., Dubuque,T., Favello,A., Gish,W., Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N., Mardis,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Meg,J., Treviskis,E., Underwood,K., Wohlmann,P., Waterston,R., Wilson,R. and Marra,M.  
Generation and analysis of 280,000 human expressed sequence tags  
Genome Res. 6 (9), 807-828 (1996)  
97044478

## COMMENT

Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu  
Insert Size: 908  
High quality sequence stops: 309 Source: IMAGE Consortium, LINT This clone is available royalty-free through LINT; contact the IMAGE Consortium ([info@image.lnlnl.gov](mailto:info@image.lnlnl.gov)) for further information.  
Insert length: 908 Std Error: 0.00



	Df	52	AAGCAATGACGTGAATTACTTATCT	T	851 bp DNA linear	GSS 19-JUL-1999
RESULT 14	Locus	A0748718/c				
DEFINITION			HS 5540 A2 B03 SP6 RPlCl-11 Human Male BAC library Homo sapiens genomic clone Plate=1116 Col=6 Row=C,		DNA sequence.	
ACCESSION		AQ748718				
VERSION		AQ748718				
KEYWORDS	SOURCE	human.				
ORGANISM		Homo sapiens				
REFERENCE		Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
AUTHORS		Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 851)				
		Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D. and Hood,L..				
TITLE		Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome				
JOURNAL MEDLINE		Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999) 99380589				
COMMENT		Contact: Mahairas GG, Wallace JC, Hood L High Throughput Sequencing Center University of Washington 401 Queen Anne Avenue North, Seattle, WA 98109, USA Tel.: (206) 616-3618 Fax: (206) 616-3887 Email: jwallace@u.washington.edu Clones are derived from the human BAC library RPlCl-11. For BAC library availability, please contact Pietter de Jong ( <a href="#">pietder@joe.med.buffalo.edu</a> ). Clones may be purchased from BACPac Resources ( <a href="#">http://bacpac.med.buffalo.edu/ordering_bac.htm</a> ) or from Resear h Genetics ( <a href="#">info@gresgen.com</a> ). BAC end Web Server: <a href="#">http://www.htsc.washington.edu</a> Plate: 1116 row: C column: 6 Seq primer: SP6 Class: BAC ends High quality sequence stop: 851.				
FEATURES	source	Location/Qualifiers  1..851 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="Plate=1116 Col=6 Row=C" /_clone_id="RPlCl-11 Human Male BAC Library" /sex="male" /note="Vector: pBACE3.6; Site_1: EcoRI; Site_2: EcoRI; Male blood DNA was isolated from one randomly chosen donor and partially digested with a combination of EcoRI and EcoRI Methylase. Size selected DNA was cloned into the pBACE3.6 vector at EcoRI sites"				
BASE COUNT		220 a                 220 g                 245 t                 1 others				
ORIGIN		188 c                 197 g                 245 t                 1 others				
Query Match		50.0%; Score 48; DB 17; Length 851;				
Best Local Similarity		79.5%; Pred. No. 4.9e-05;				
Matches 70;	Conservative	0; Mismatches 15; Indels 3; Gaps 1;				
OY	12	C TCCCGAAGAAAGTAATAATTATACACCACCTTTACACTTAGACCTCTTGTT---AGAA 68				
Db	156	CACCGAGAGAAGTAATAATTATTCACAATTATCCTGCAGCTTCTGTGTAGAGAGA 97				
OY	69	AAGCAAATGGAGTGAATGCCCATAGT 96				
Db	96	AAGCAAATGGAGTGAATCACCTTATGT 69				
RESULT 15						
LOCUS	A0000365/c					
DEFINITION		CIT-HSP 2283L7.TR CIT-HSP Homo sapiens genomic clone 2283L7, DNA sequence.				

```

ACCESSION      A0000365
VERSION        A0000365.1
KEYWORDS       GI:3027804
SOURCE         human.
ORGANISM       Homo sapiens
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE      1 (bases 1 to 632)
               Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K.,
               Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H.,
               Simon,M. and Venter,J.C.
               Use of a random BAC End Sequence Database for Sequence-Ready Map
               Building (1998)
TITLE          Unpublished (1998)
JOURNAL        Other-GSSs: CIT-HSP-2283L7.TF
COMMENT        Contact: Mark Adams
               Department of Eukaryotic Genomics
               The Institute for Genomic Research
               9712 Medical Center Dr., Rockville, MD 20850, USA
               Tel.: 301 838 0200
               Fax: 301 838 0208
               Email: mdadams@tigr.org
               Clones are available from Research Genetics (info@resgen.com). BAC
               end search page:
               http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html
               Seq primer: M13 Reverse
               Class: BAC ends.
FEATURES       Location/Qualifiers
               source          1..632
                                /organism="Homo sapiens"
                                /db_xref="GDB:714807a"
                                /db_xref="taxon:9606"
                                /clone="2283L7"
                                /clone_11b="CIT-HSP"
                                /sex="Male"
                                /cell_type="Sperm"
                                /note="Vector: pBeloBAC11; Site_1: HindIII; Site_2:
                                HindIII"
BASE COUNT     154 a      143 c      151 g      184 t
ORIGIN
Query Match    49.6%; Score 47.6; DB 17; Length 632;
Best Local Similarity 76.6%; Pred. No. 6e-05;
Matches 72; Conservative 0; Mismatches 19; Indels 3; Gaps 1
QY      6 CTGACGACTCCTGAGGAGATATTAATTATACACCATCTTACAGCTAGACCTCTTTGTA 65
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      155 CTGGCCCCCTCAGGAGAGTATTAATTATTAAGCCCTCTTACACATGATCTTTCTGTA 96
QY      66 GAA---AAGGCAATGAGAGTGAAGTCCCTAAGT 96
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB      95 AACGAGAGGATAAATGAGATGAAGTCCCTTATGT 62

```





US-09-078-294-3

Query Match	46.7%	Score	44.8;	DB	4;	length	80595;
Best Local Similarity	77.3%	Pred. No.	2e-06;				
Matches	68;	Conservative	0;	Mismatches	17;	Indels	3;
						Gaps	1

QY	12	CTCTGAGGCAATATTAATTTTAAACACACTTTACACTGACCTCTTTTGGAGAA---	68
Db	56212	CACCTGAGGAAAGTATAAATTACATACATCTCCACCTTTCTCTGTACAGAG	56271
QY	69	AAGCAAAATGAGATGAGAGTCCATAAGT	96
Db	56272	AAGCAAAATGAGATGAGAAATACCTTATAGT	56299

RESULT 3  
US-08-691-563C-60  
; Sequence 60, Application US/08691563C  
; Patent No. 6001087

```

: GENERAL INFORMATION:
: APPLICANT: Herve PERRON
: APPLICANT: Frederic BESEME
: APPLICANT: Frederic BEDIN
: APPLICANT: Glaucia PARANHOS-BACCALA
: APPLICANT: Florence KOMURIAN-PRADEL
: APPLICANT: Colette JOLIVET
: APPLICANT: Bernard MANDRAND
: TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
: TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTIC
: TITLE OF INVENTION: THERAPEUTIC PURPOSES
: NUMBER OF SEQUENCES: 92
: CORRESPONDENCE ADDRESS:

```

Query Match	41.0%;	Score 39.4;	DB 3;	Length 2503;
Best Local Similarity	76.6%;	Pred. No. 5.9e-05;		
Matches 59; Conservative	2;	Mismatches 13;	Indels 3;	Gaps 1;

QY 12 CTCCTGAGGGAAATATAAATTATAACACATCTTAAGCTAAGCCTTTTGTAGAAA--- 68  
| | | | | : | | | | |  
Db 1496 CACCTTGAGGGGAACCAAAAAATTACAATAATATCTTAAGCYATCTTTCTTAGAGAGG 155  
| | | | |  
QY 69 AAGGCMAATGAGATGAA 85  
| | | | |

Db 1556 AAGCAAATGGAGTGAA 1572

RESULT 4  
US-09-149-476-30  
; Sequence 30, Application US/09149476

GENERAL INFORMATION:	
APPLICANT:	Rosen et al.
TITLE OF INVENTION:	166 Human Secreted proteins
FILE REFERENCE:	P200201
CURRENT APPLICATION NUMBER:	US/09/149,476
CURRENT FILING DATE:	1998-09-08
EARLIER APPLICATION NUMBER:	PCT/US98/04493
EARLIER FILING DATE:	1998-03-06
EARLIER APPLICATION NUMBER:	60/040,162
EARLIER FILING DATE:	1997-03-07
EARLIER APPLICATION NUMBER:	60/040,333
EARLIER FILING DATE:	1997-03-07
EARLIER APPLICATION NUMBER:	60/038,621
EARLIER FILING DATE:	1997-03-07
EARLIER APPLICATION NUMBER:	60/040,626
EARLIER FILING DATE:	1997-03-07
EARLIER APPLICATION NUMBER:	60/040,334
EARLIER FILING DATE:	1997-03-07
EARLIER APPLICATION NUMBER:	60/040,336
EARLIER FILING DATE:	1997-03-07
EARLIER APPLICATION NUMBER:	60/040,163
EARLIER FILING DATE:	1997-03-07
EARLIER APPLICATION NUMBER:	60/047,600
EARLIER FILING DATE:	1997-05-23
EARLIER APPLICATION NUMBER:	60/047,615
EARLIER FILING DATE:	1997-05-23
EARLIER APPLICATION NUMBER:	60/047,597
EARLIER FILING DATE:	1997-05-23
EARLIER APPLICATION NUMBER:	60/047,502
EARLIER FILING DATE:	1997-05-23
EARLIER APPLICATION NUMBER:	60/047,633
EARLIER FILING DATE:	1997-05-23
EARLIER APPLICATION NUMBER:	60/047,583
EARLIER FILING DATE:	1997-05-23
EARLIER APPLICATION NUMBER:	60/047,617
EARLIER FILING DATE:	1997-05-23
EARLIER APPLICATION NUMBER:	60/047,581
EARLIER FILING DATE:	1997-05-23
EARLIER APPLICATION NUMBER:	60/047,584
EARLIER FILING DATE:	1997-05-23
EARLIER APPLICATION NUMBER:	60/047,500
EARLIER FILING DATE:	1997-05-23
EARLIER APPLICATION NUMBER:	60/047,587
EARLIER FILING DATE:	1997-05-23
EARLIER APPLICATION NUMBER:	60/047,492
EARLIER FILING DATE:	1997-05-23
EARLIER APPLICATION NUMBER:	60/047,598
EARLIER FILING DATE:	1997-05-23
EARLIER APPLICATION NUMBER:	60/047,613
EARLIER FILING DATE:	1997-05-23
EARLIER APPLICATION NUMBER:	60/047,582
EARLIER FILING DATE:	1997-05-23
EARLIER APPLICATION NUMBER:	60/047,596
EARLIER FILING DATE:	1997-05-23
EARLIER APPLICATION NUMBER:	60/047,612
EARLIER FILING DATE:	1997-05-23
EARLIER APPLICATION NUMBER:	60/047,632
EARLIER FILING DATE:	1997-05-23
EARLIER APPLICATION NUMBER:	60/047,601
EARLIER FILING DATE:	1997-05-23



EARLIER APPLICATION NUMBER:	60/043, 580
EARLIER FILING DATE:	1997-04-11
EARLIER APPLICATION NUMBER:	60/043, 568
EARLIER FILING DATE:	1997-04-11
EARLIER APPLICATION NUMBER:	60/043, 568
EARLIER FILING DATE:	1997-04-11
EARLIER APPLICATION NUMBER:	60/043, 569
EARLIER FILING DATE:	1997-04-11
EARLIER APPLICATION NUMBER:	60/043, 311
EARLIER FILING DATE:	1997-04-11
EARLIER APPLICATION NUMBER:	60/043, 671
EARLIER FILING DATE:	1997-04-11
EARLIER APPLICATION NUMBER:	60/043, 674
EARLIER FILING DATE:	1997-04-11
EARLIER APPLICATION NUMBER:	60/043, 669
EARLIER FILING DATE:	1997-04-11
EARLIER APPLICATION NUMBER:	60/043, 312
EARLIER FILING DATE:	1997-04-11
EARLIER APPLICATION NUMBER:	60/043, 313
EARLIER FILING DATE:	1997-04-11
EARLIER APPLICATION NUMBER:	60/043, 672
EARLIER FILING DATE:	1997-04-11
EARLIER APPLICATION NUMBER:	60/043, 315
EARLIER FILING DATE:	1997-04-11
EARLIER APPLICATION NUMBER:	60/048, 974
EARLIER FILING DATE:	1997-06-06
EARLIER APPLICATION NUMBER:	60/056, 886
EARLIER FILING DATE:	1997-08-22
EARLIER APPLICATION NUMBER:	60/056, 877
EARLIER FILING DATE:	1997-08-22
EARLIER APPLICATION NUMBER:	60/056, 889
EARLIER FILING DATE:	1997-08-22
EARLIER APPLICATION NUMBER:	60/056, 893
EARLIER FILING DATE:	1997-08-22
EARLIER APPLICATION NUMBER:	60/056, 630
EARLIER FILING DATE:	1997-08-22
EARLIER APPLICATION NUMBER:	60/056, 878
EARLIER FILING DATE:	1997-08-22
EARLIER APPLICATION NUMBER:	60/056, 662
EARLIER FILING DATE:	1997-08-22
EARLIER APPLICATION NUMBER:	60/056, 872
EARLIER FILING DATE:	1997-08-22
EARLIER APPLICATION NUMBER:	60/056, 888
EARLIER FILING DATE:	1997-08-22
EARLIER APPLICATION NUMBER:	60/056, 637
EARLIER FILING DATE:	1997-08-22
EARLIER APPLICATION NUMBER:	60/056, 903
EARLIER FILING DATE:	1997-08-22
EARLIER APPLICATION NUMBER:	60/056, 888
EARLIER FILING DATE:	1997-08-22
EARLIER APPLICATION NUMBER:	60/056, 879
EARLIER FILING DATE:	1997-08-22
EARLIER APPLICATION NUMBER:	60/056, 880
EARLIER FILING DATE:	1997-08-22
EARLIER APPLICATION NUMBER:	60/056, 874
EARLIER FILING DATE:	1997-08-22
EARLIER APPLICATION NUMBER:	60/056, 910
EARLIER FILING DATE:	1997-08-22
EARLIER APPLICATION NUMBER:	60/056, 864
EARLIER FILING DATE:	1997-08-22
EARLIER APPLICATION NUMBER:	60/056, 631
EARLIER FILING DATE:	1997-08-22
EARLIER APPLICATION NUMBER:	60/056, 845
EARLIER FILING DATE:	1997-08-22
EARLIER APPLICATION NUMBER:	60/056, 892
EARLIER FILING DATE:	1997-08-22
EARLIER APPLICATION NUMBER:	60/057, 761

EARLIER	FILING DATE:	1997-08-22
EARLIER	APPLICATION NUMBER:	60/047,595
EARLIER	FILING DATE:	1997-05-23
EARLIER	APPLICATION NUMBER:	60/047,599
EARLIER	FILING DATE:	1997-05-23
EARLIER	APPLICATION NUMBER:	60/047,588
EARLIER	FILING DATE:	1997-05-23
EARLIER	APPLICATION NUMBER:	60/047,585
EARLIER	FILING DATE:	1997-05-23
EARLIER	APPLICATION NUMBER:	60/047,586
EARLIER	FILING DATE:	1997-05-23
EARLIER	APPLICATION NUMBER:	60/047,590
EARLIER	FILING DATE:	1997-05-23
EARLIER	APPLICATION NUMBER:	60/047,594
EARLIER	FILING DATE:	1997-05-23
EARLIER	APPLICATION NUMBER:	60/047,589
EARLIER	FILING DATE:	1997-05-23
EARLIER	APPLICATION NUMBER:	60/047,593
EARLIER	FILING DATE:	1997-05-23
EARLIER	APPLICATION NUMBER:	60/047,614
EARLIER	FILING DATE:	1997-05-23
EARLIER	APPLICATION NUMBER:	60/043,578
EARLIER	FILING DATE:	1997-04-11
EARLIER	APPLICATION NUMBER:	60/043,576
EARLIER	FILING DATE:	1997-04-11
EARLIER	APPLICATION NUMBER:	60/047,501
EARLIER	FILING DATE:	1997-05-23
EARLIER	APPLICATION NUMBER:	60/043,670
EARLIER	FILING DATE:	1997-04-11
EARLIER	APPLICATION NUMBER:	60/056,632
EARLIER	FILING DATE:	1997-08-22
EARLIER	APPLICATION NUMBER:	60/056,664
EARLIER	FILING DATE:	1997-08-22
EARLIER	APPLICATION NUMBER:	60/056,876
EARLIER	FILING DATE:	1997-08-22
EARLIER	APPLICATION NUMBER:	60/056,881
EARLIER	FILING DATE:	1997-08-22
EARLIER	APPLICATION NUMBER:	60/056,909
EARLIER	FILING DATE:	1997-08-22
EARLIER	APPLICATION NUMBER:	60/056,875
EARLIER	FILING DATE:	1997-08-22
EARLIER	APPLICATION NUMBER:	60/056,862
EARLIER	FILING DATE:	1997-08-22
EARLIER	APPLICATION NUMBER:	60/056,887
EARLIER	FILING DATE:	1997-08-22
EARLIER	APPLICATION NUMBER:	60/056,908
EARLIER	FILING DATE:	1997-08-22
EARLIER	APPLICATION NUMBER:	60/048,964
EARLIER	FILING DATE:	1997-06-06
EARLIER	APPLICATION NUMBER:	60/057,650
EARLIER	FILING DATE:	1997-09-05
EARLIER	APPLICATION NUMBER:	60/056,884
EARLIER	FILING DATE:	1997-08-22
EARLIER	APPLICATION NUMBER:	60/057,669
EARLIER	FILING DATE:	1997-09-05
EARLIER	APPLICATION NUMBER:	60/049,610
EARLIER	FILING DATE:	1997-06-13
EARLIER	APPLICATION NUMBER:	60/061,060
EARLIER	FILING DATE:	1997-10-02

Query Match	29.6%	Score 28.4	DB 4	Length 434
Best Local Similarity	60.3%	Pred. No. 0.3		
Matches	41	Conservative	3	Mismatches 24; Indels 0; Gaps 0
QY	28	AAATTTAAACACACTTACAGCTGACTCTCTTTGTGAAAAGCAATGGAGTGA	87	
Db	190	AAATGTTAAATATGTTCCATCTGACATGTTATTACCAARAAAARAGATTAATTC	249	
QY	88	GCCATTAAG	95	
Db	250	TACATCAG	257	

RESULT 5  
US-08-378-313-20/c  
; Sequence 20, Application US/08378313  
; Patent No. 6207881  
; GENERAL INFORMATION:  
; APPLICANT: THEOLOGIS, ATHANASIOS  
; APPLICANT: SATO, TAKAHIDO  
; TITLE OF INVENTION: CONTROL OF FRUIT RIPENING THROUGH  
; TITLE OF INVENTION: GENETIC CONTROL OF ACC SYNTHASE SYNTHESIS  
; NUMBER OF SEQUENCES: 34  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: MORRISON & FOERSTER  
; STREET: 755 Page Mill Road  
; CITY: Palo Alto  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94304-1018  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/378,313  
; FILING DATE:  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/862,493  
; FILING DATE: 02-APR-1992  
; ATTORNEY/AGENT INFORMATION:  
; NAME: MORASHIGE, KATE H.  
; REGISTRATION NUMBER: 29,959  
; REFERENCE/DOCKET NUMBER: 29190-20002.20  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 856-5600  
; TELEFAX: (415) 494-0792  
; TELEX: 706141  
; INFORMATION FOR SEQ ID NO: 20:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 9060 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: join(2704..2880, 2968..3099, 3183..3344, 3810  
; LOCATION: ..4376, 4463..4903)  
; US-08-378-313-20  
  
Query Match 28.1%; Score 27; DB 4; Length 9060;  
Best Local Similarity 66.1%; Pred. No. 2.4;  
Matches 39; Conservative 0; Mismatches 20; Indels 0; Gaps 0;  
  
QY 19 GGGAGCTAATATTAACACCATCTTACAGCTAGACCTCTTTGTGAAAGGCAAT 77  
DB 7306 GGGAGCTAATATTAATATCAACACCATCTTACAGCTTGTGTTAAAGTCAAT 7248  
  
RESULT 6  
US-08-252-626A-1/c  
; Sequence 1, Application US/08252626A.  
; Patent No. 5385269  
; GENERAL INFORMATION:  
; APPLICANT: Earp, Henry S.  
; APPLICANT: Graham, Douglas K.  
; APPLICANT: Dawson, Thomas L.  
; APPLICANT: Mulaney, David U.  
; APPLICANT: Snodgrass, Hiram R.  
; TITLE OF INVENTION: Isolated DNA Encoding C-MER  
; TITLE OF INVENTION: Protooncogene  
; NUMBER OF SEQUENCES: 9  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Kenneth D. Sibley

STREET: P.O. Drawer 34009  
CITY: Charlotte  
STATE: No. 5585269th Carolina  
COUNTRY: USA  
ZIP: 28234  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/252,626A  
; FILING DATE: 02-JUN-1994  
; CLASSIFICATION: 536  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Sibley, Kenneth D.  
; REGISTRATION NUMBER: 31,665  
; REFERENCE/DOCKET NUMBER: 5470-81  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (919) 881-3140  
; TELEFAX: (919) 881-3175  
; TELEX: 575102  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3635 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 138..3137  
; US-08-252-626A-1  
  
Query Match 27.7%; Score 26.6; DB 1; Length 3635;  
Best Local Similarity 60.3%; Pred. No. 2.5;  
Matches 44; Conservative 0; Mismatches 29; Indels 0; Gaps 0;  
  
QY 24 GTATTAATTAATTAACCATCTTACAGCTAGACCTCTTTGTGAAAGGCAATGAGTG 83  
DB 3445 GTACAAAGCTTAATATCAAGGAGCGCTGAAGCTTAATTAATTAATGACATATGCAAGTAAAGTG 3386  
  
QY 84 AAGTGCCATAAGT 96  
DB 3385 AATGGAATAGT 3373  
  
RESULT 7  
US-08-286-819A-24/c  
; Sequence 24, Application US/08286819A  
; Patent No. 5871910  
; GENERAL INFORMATION:  
; APPLICANT: ARTHUR, MICHEL  
; APPLICANT: DURTA-MALEN, SYLVIE  
; APPLICANT: MOLINAS, CATHERINE  
; APPLICANT: COURVALIN, PATRICE  
; TITLE OF INVENTION: POLYPEPTIDES IMPLICATED IN THE  
; TITLE OF INVENTION: EXPRESSION OF RESISTANCE TO GLYCOPROTEINS. IN PARTICULAR  
; TITLE OF INVENTION: IN GRAM-POSITIVE BACTERIA, NUCLEOTIDE SEQUENCE CODING FOR  
; THESE POLYPEPTIDES AND USE FOR DIAGNOSIS  
; NUMBER OF SEQUENCES: 54  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,  
; ADDRESSEE: P.C.  
; STREET: 1755 S. Jefferson Davis Highway, Suite 400  
; CITY: Arlington  
; STATE: Virginia  
; COUNTRY: U.S.A.  
; ZIP: 22202  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/286,819A  
FILING DATE: 05-AUG-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/174,682  
FILING DATE: 28-DEC-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/917,146  
FILING DATE: 10-AUG-1992  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/FR/91/00855  
FILING DATE: 29-OCT-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: FR 9013579  
FILING DATE: 31-OCT-1990  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: OBLON, NO. 587191dman F.  
REGISTRATION NUMBER: 24,618  
REFERENCE/DOCKET NUMBER: 660-060-0 PCT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 413-3000  
TELEFAX: (703) 413-2220  
TELEX: 248855 OPAT UR  
INFORMATION FOR SEQ ID NO: 24:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 483 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: unknown  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..483  
US-08-286-819A-24

Query Match 26.5%; Score 25.4; DB 2; Length 483;  
Best Local Similarity 61.2%; Pred. No. 3.6;  
Matches 41; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

OY 21 GAAGTAAATATATACACCATCTTACAGCTAGACCTCTTTGTACAAAAGCAATGGA 80  
DB 188 GGATTAAGATTATACATTACTATCACTCTCTGAAATCCACGATAGCAATGGA 129

OY 81 GTGAAGT 87  
DB 128 GTCAAGT 122

RESULT 8  
US-08-980-357-24/c  
Sequence 24, Application US/08980357  
Patent No. 6013508  
GENERAL INFORMATION:  
APPLICANT: ARTHUR, MICHEL  
APPLICANT: DUKTA-MALEN, SYLVIE  
APPLICANT: MOLINAS, CATHERINE  
APPLICANT: COURVALIN, PATRICE  
TITLE OF INVENTION: POLYPEPTIDES IMPLICATED IN THE  
TITLE OF INVENTION: EXPRESSION OF RESISTANCE TO GLYCOPROTEINS, IN PARTICULAR  
TITLE OF INVENTION: IN GRAM-POSITIVE BACTERIA, NUCLEOTIDE SEQUENCE CODING FOR  
NUMBER OF SEQUENCES: 54  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT,  
ADDRESSEE: P.C.  
STREET: 1755 S. Jefferson Davis Highway, Suite 400  
CITY: Arlington  
STATE: Virginia

COUNTRY: U.S.A.  
ZIP: 22202  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/980,357  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/286,819  
FILING DATE: 05-AUG-1994  
APPLICATION NUMBER: US 08/174,682  
FILING DATE: 28-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/917,146  
FILING DATE: 10-AUG-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/FR/91/00855  
FILING DATE: 29-OCT-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: FR 9013579  
FILING DATE: 31-OCT-1990  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: OBLON, NO. 601350dman F.  
REGISTRATION NUMBER: 24,618  
REFERENCE/DOCKET NUMBER: 660-060-0 PCT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 413-3000  
TELEFAX: (703) 413-2220  
TELEX: 248855 OPAT UR  
INFORMATION FOR SEQ ID NO: 24:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 483 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: unknown  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1..483  
US-08-980-357-24

Query Match 26.5%; Score 25.4; DB 3; Length 483;  
Best Local Similarity 61.2%; Pred. No. 3.6;  
Matches 41; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

OY 21 GAAGTAAATATATACACCATCTTACAGCTAGACCTCTTTGTACAAAAGCAATGGA 80  
DB 188 GGATTAAGATTATACATTACTATCACTCTCTGAAATCCACGATAGCAATGGA 129

OY 81 GTGAAGT 87  
DB 128 GTCAAGT 122

RESULT 9  
US-08-286-819A-16/c  
Sequence 16, Application US/08286819A  
Patent No. 5871910  
GENERAL INFORMATION:  
APPLICANT: ARTHUR, MICHEL  
APPLICANT: DUKTA-MALEN, SYLVIE  
APPLICANT: MOLINAS, CATHERINE  
APPLICANT: COURVALIN, PATRICE  
TITLE OF INVENTION: POLYPEPTIDES IMPLICATED IN THE  
TITLE OF INVENTION: EXPRESSION OF RESISTANCE TO GLYCOPROTEINS, IN PARTICULAR  
TITLE OF INVENTION: IN GRAM-POSITIVE BACTERIA, NUCLEOTIDE SEQUENCE CODING FOR  
NUMBER OF SEQUENCES: 54  
CORRESPONDENCE ADDRESS:

ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT,  
ADDRESSEE: P.C. S. Jefferson Davis Highway, Suite 400  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/286,819A  
FILING DATE: 05-AUG-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/174,682  
FILING DATE: 28-DEC-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/917,146  
FILING DATE: 10-AUG-1992  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/FR/91/00855  
FILING DATE: 29-OCT-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: FR 9013579  
FILING DATE: 31-OCT-1990  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Oblon, No. 5871910man F.  
REGISTRATION NUMBER: 24,618  
REFERENCE/DOCKET NUMBER: 660-060-0 PCT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 413-3000  
TELEFAX: (703) 413-2220  
TELEX: 248855 OPAT UR  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 10851 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA (genomic)  
US-08-286-819A-16  
Query Match 26.5%; Score 25.4; DB 2; Length 10851;  
Best Local Similarity 61.2%; Pred. No. 9.4;  
Matches 41; Conservative 0; Mismatches 26; Indels 0; Gaps 0;  
QY 21 GAAGTAAATTATACACCATCTTACAGCTAGACCTTTTGTAGAAAAGCAATGGA 80  
DB 10303 GGAATTAAGATTATACATATCATCATCTCTGAAATTCACAGTAGCAATGGA 10244  
QY 81 GTGAAGT 87  
DB 10243 GTCAAGT 10237  
RESULT 10  
US-08-980-357-16/C  
; Sequence 16, Application US/08980357  
; Patent No. 6013508  
; GENERAL INFORMATION:  
; APPLICANT: ARTHUR, MICHEL  
; APPLICANT: DUKTA-MALEN, SYLVIE  
; APPLICANT: MOLINAS, CATHERINE  
; APPLICANT: COURVALIN, PATRICE  
; TITLE OF INVENTION: POLYPEPTIDES IMPLICATED IN THE  
; TITLE OF INVENTION: EXPRESSION OF RESISTANCE TO GLYCOPROTEINS, IN PARTICULAR  
; TITLE OF INVENTION: IN GRAM-POSITIVE BACTERIA, NUCLEOTIDE SEQUENCE CODING FOR

TITLE OF INVENTION: THESE POLYPEPTIDES AND USE FOR DIAGNOSIS  
NUMBER OF SEQUENCES: 54  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: OBLON, SPIVAK, MCLELLAND, MAIER & NEUSTADT,  
ADDRESSEE: P.C.  
CITY: Arlington  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22202  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/980,357  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/286,819  
FILING DATE: 05-AUG-1994  
APPLICATION NUMBER: US 08/174,682  
FILING DATE: 28-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/917,146  
FILING DATE: 10-AUG-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/FR/91/00855  
FILING DATE: 29-OCT-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: FR 9013579  
FILING DATE: 31-OCT-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Oblon, No. 6013508man F.  
REGISTRATION NUMBER: 24,618  
REFERENCE/DOCKET NUMBER: 660-060-0 PCT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 413-3000  
TELEFAX: (703) 413-2220  
TELEX: 248855 OPAT UR  
INFORMATION FOR SEQ ID NO: 16:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 10851 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA (genomic)  
US-08-980-357-16  
Query Match 26.5%; Score 25.4; DB 3; Length 10851;  
Best Local Similarity 61.2%; Pred. No. 9.4;  
Matches 41; Conservative 0; Mismatches 26; Indels 0; Gaps 0;  
QY 21 GAAGTAAATTATACACCATCTTACAGCTAGACCTTTTGTAGAAAAGCAATGGA 80  
DB 10303 GGAATTAAGATTATACATATCATCATCTCTGAAATTCACAGTAGCAATGGA 10244  
QY 81 GTGAAGT 87  
DB 10243 GTCAAGT 10237  
RESULT 11  
US-08-858-207A-43  
; Sequence 43, Application US/08858207A  
; Patent No. 6348328  
; GENERAL INFORMATION:  
; APPLICANT: Black, Michael  
; APPLICANT: Hodgson, John  
; APPLICANT: Knowles, David  
; APPLICANT: Nicholas, Richard  
; APPLICANT: Stodola, Robert

TITLE OF INVENTION: No. 6348328e1 Compounds  
NUMBER OF SEQUENCES: 552  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SmithKline Beecham Corporation  
STREET: 709 Swedeland Road  
CITY: King of Prussia  
STATE: PA  
COUNTRY: USA  
ZIP: 19406-0939  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/858,207A  
FILING DATE: 09-MAY-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/017670  
FILING DATE: 14-MAY-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Gimmil, Edward R.  
REGISTRATION NUMBER: 38,891  
REFERENCE/DOCKET NUMBER: P50475  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 610-270-4478  
TELEFAX: 610-270-5090  
TELEX:  
INFORMATION FOR SEQ ID NO: 43:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 874 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-858-207A-43  
Query Match 26.2%; Score 25.2; DB 4; Length 874;  
Best Local Similarity 57.7%; Pred. No. 5.1;  
Matches 45; Conservative 0; Mismatches 33; Indels 0; Gaps 0;  
QY 19 GGAAGTATTAATTAACACCATCTTACAGCTAGTCTTTTGTAGAAAGCAAAAG 78  
DB 218 GAGAAGATTAAGATACACCATGACAGCTTAAGCTTATCTTAAGGCGCTACT 277  
QY 79 GAGTGAAGTGCCTAAGT 96  
DB 278 GATGACCTTAGAATAGT 295  
RESULT 12  
US-08-117-083-59  
Sequence 59, Application US/08117083  
Patent No. 5719054  
GENERAL INFORMATION:  
APPLICANT: Bournell, Michael E.  
APPLICANT: Ingils, Stephen C.  
APPLICANT: Munro, Alan J.  
TITLE OF INVENTION: Recombinant Virus Vectors Encoding Human  
NUMBER OF SEQUENCES: 70  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Walter H. Dreyer  
STREET: 4 Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: CA  
COUNTRY: USA  
ZIP: 94111  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC Compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/117,083  
FILING DATE: 10-SEP-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Dreyer, Walter H.  
REGISTRATION NUMBER: 24,190  
REFERENCE/DOCKET NUMBER: A-58783  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415-781-1989  
TELEFAX: 415-398-3249  
TELEX: 910 277299  
INFORMATION FOR SEQ ID NO: 59:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1250 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
US-08-117-083-59  
Query Match 26.2%; Score 25.2; DB 1; Length 1250;  
Best Local Similarity 62.9%; Pred. No. 5.7;  
Matches 39; Conservative 0; Mismatches 23; Indels 0; Gaps 0;  
QY 22 AAGTATTAATTAACACCATCTTACAGCTAGTCTTTTGTAGAAAGCAAAAGGAG 81  
DB 1186 AATATTAATTAATTAATTAATCTTACAGCGCCCATCATGAAATACGACGATTAATGAT 1245  
QY 82 TG 83  
DB 1246 TG 1247  
RESULT 13  
US-09-157-257-7  
Sequence 7, Application US/09157257  
Patent No. 6375954  
GENERAL INFORMATION:  
APPLICANT: Dutta, Sukanta K.  
APPLICANT: Biswas, Biswajit  
APPLICANT: Vemulapalli, Ramesh  
TITLE OF INVENTION: A SIZE-VARIABLE STRAIN-SPECIFIC PROTECTIVE ANTIGEN FOR  
TITLE OF INVENTION: POTOMAC HORSE FEVER  
FILE REFERENCE: 8172-9016  
CURRENT APPLICATION NUMBER: US/09/157,257  
CURRENT FILING DATE: 1998-09-18  
EARLIER APPLICATION NUMBER: 60/059,252  
EARLIER FILING DATE: 1997-09-18  
NUMBER OF SEQ ID NOS: 48  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 7  
LENGTH: 1791  
TYPE: DNA  
ORGANISM: Ehrlichia risticii  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (175)..(1677)  
US-09-157-257-7  
Query Match 26.2%; Score 25.2; DB 4; Length 1791;  
Best Local Similarity 60.0%; Pred. No. 6.4;  
Matches 42; Conservative 0; Mismatches 28; Indels 0; Gaps 0;  
QY 19 GGAAGTATTAATTAACACCATCTTACAGCTAGTCTTTTGTAGAAAGCAAAAG 78  
DB 496 GGTAAAGATATATTAAACACATCTACACAGATAGCAGTAAAGCTTTAAAGAGCTGATA 555  
QY 79 GAGTGAAGTG 88  
DB 556 GGAAGTAGCG 565

```
RESULT 14
US-09-268-364-15
; Sequence 15, Application US/09268364A
; Patent No. 6204063
; GENERAL INFORMATION:
; APPLICANT: Allen, Steve
; APPLICANT: Lee, Jian Ming
; APPLICANT: Lightner, Jonathan
; TITLE OF INVENTION: PLANT GLYCOLYSIS AND RESPIRATION ENZYMES
; FILE REFERENCE: BB-1154
; CURRENT APPLICATION NUMBER: US/09/268,364A
; EARLIER FILING DATE: 1998-03-15
; EARLIER APPLICATION NUMBER: 60/079,387
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 15
; LENGTH: 857
; TYPE: DNA
; ORGANISM: Oryza sativa
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (11)
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (115)
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (744)
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (767)
; FEATURE:
; NAME/KEY: (794)..(795)
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (806)
; US-09-268-364-15

Query Match
Best Local Similarity 25.8%; Score 24.8; DB 4; Length 857;
Matches 32; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 40 CATCTACAGCTGACCTTTTGTAGAAAAGCCAAATGAGATG 83
Db 295 CATCTACACCTGACCTATTTGCTTACAAAGCATGTGTAGAG 338

RESULT 15
US-08-755-587-5/C
; Sequence 5, Application US/08755587
; Patent No. 6045997
; GENERAL INFORMATION:
; APPLICANT: Futreal, Phillip A
; APPLICANT: Wooster, Richard F
; APPLICANT: Asmworth, Alan
; TITLE OF INVENTION: Materials and methods relating to the
; TITLE OF INVENTION: Identification and sequencing of the BRCA2 cancer
; TITLE OF INVENTION: susceptibility gene and uses thereof.
; NUMBER OF SEQUENCES: 222
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Bell Seltzer Park & Gibson
; STREET: 310 UCB Plaza, 3605 Glenwood Avenue, PO Drawer 31107
; CITY: Raleigh
; STATE: NC
; COUNTRY: USA
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
```

```
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/755,587
; FILING DATE: 25-NOV-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9523959.6
; FILING DATE: 23-NOV-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9525555.0
; FILING DATE: 14-DEC-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9617961.9
; FILING DATE: 28-AUG-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Kenneth D Sibley
; REGISTRATION NUMBER: 31,665
; REFERENCE/DOCKET NUMBER: 5405-135
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 243 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-755-587-5

Query Match
Best Local Similarity 25.6%; Score 24.6; DB 3; Length 243;
Matches 42; Conservative 0; Mismatches 29; Indels 0; Gaps 0;

QY 7 TGGCAGCTCGTGGAGGAGTATTAATTATACACCATCTTACAGCTAGACCTCTTTGTAG 66
Db 119 TGGATCACCTGAGTGACAGATATTTATATACATACCTTATAGAGGAGAGATATTAAT 60

QY 67 AAAAGGCAAT 77
Db 59 AAAAGCTAAT 49
```

Search completed: May 2, 2003, 12:49:12  
Job time : 78.4 secs

GenCore version 5.1.5  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 2, 2003, 12:25:30 ; Search time 63.8897 seconds  
(without alignments)  
1775.493 Million cell updates/sec

Title: US-09-719-554-3\_COPY\_3065\_3160

Perfect score: 96  
Sequence: 1 gccgcctgcacctcctgagc.....tgagtgtaagtgccataagt 96

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 746064 seqs, 590810554 residues

Total number of hits satisfying chosen parameters: 1492128

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

Published\_Applications\_NA:\*

- 1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*
- 2: /cgn2\_6/ptodata/1/pubpna/PC1\_NEM\_PUB.seq:\*
- 3: /cgn2\_6/ptodata/1/pubpna/US06\_NEM\_PUB.seq:\*
- 4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:\*
- 5: /cgn2\_6/ptodata/1/pubpna/US07\_NEM\_PUB.seq:\*
- 6: /cgn2\_6/ptodata/1/pubpna/PC1S\_PUBCOMB.seq:\*
- 7: /cgn2\_6/ptodata/1/pubpna/US08\_NEM\_PUB.seq:\*
- 8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*
- 9: /cgn2\_6/ptodata/1/pubpna/US09\_NEM\_PUB.seq:\*
- 10: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq:\*
- 11: /cgn2\_6/ptodata/1/pubpna/US10\_NEM\_PUB.seq:\*
- 12: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq:\*
- 13: /cgn2\_6/ptodata/1/pubpna/US60\_NEM\_PUB.seq:\*
- 14: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	67.4	70.2	764	7	US-08-979-847-126
2	67.4	70.2	800	7	US-08-979-847-127
3	64.6	67.3	1511	7	US-08-979-847-117
4	64	66.7	1359	9	US-10-125-237-2
5	64	66.7	1359	9	US-10-105-891-2
6	46.4	48.3	570	10	US-09-864-761-12831
7	44.8	46.7	584	9	US-09-924-400-316
8	44.8	46.7	800	9	US-09-810-936-317
9	44.8	46.7	800	9	US-09-924-400-307
10	44.8	46.7	800	10	US-09-825-301-11
11	44.8	46.7	800	10	US-08-910-936-307
12	44.8	46.7	800	10	US-09-429-755-307
13	44.8	46.7	829	9	US-09-924-400-317
14	44.8	46.7	829	10	US-09-810-936-317
15	42.4	44.2	502	7	US-09-783-590-4203
16	39.4	41.0	2503	10	US-08-979-847-56
17	31.4	32.7	42999	10	US-09-740-029-3
18	31	32.3	565	10	US-09-864-761-16360
19	30.6	31.9	487	9	US-09-918-995-6263

C 20	30	31.2	1691139	9	US-10-067-514-1	Sequence 1, Appl
C 21	29.4	30.6	5251	9	US-10-091-504-1431	Sequence 1431, Ap
C 22	29.4	30.6	5251	10	US-09-764-869-1431	Sequence 1431, Ap
C 23	28.4	28.6	434	9	US-09-809-391-30	Sequence 30, Appl
C 24	27.8	28.0	590	9	US-10-091-572-50	Sequence 50, Appl
C 25	27.8	29.0	590	9	US-09-764-891-686	Sequence 686, App
C 26	27.8	29.0	600	9	US-10-091-572-543	Sequence 543, App
C 27	27.8	29.0	600	9	US-09-764-891-6682	Sequence 6682, App
C 28	27.6	28.8	558	10	US-09-864-761-16507	Sequence 16507, A
C 29	27.4	28.5	2732	10	US-09-925-302-178	Sequence 178, App
C 30	27	28.1	423	10	US-09-783-590-10539	Sequence 10539, A
C 31	26.6	27.7	512	10	US-09-880-107-1443	Sequence 1443, Ap
C 32	26.6	27.7	2000	9	US-09-938-842A-4348	Sequence 4348, Ap
C 33	26.6	27.7	3608	9	US-10-174-590-433	Sequence 433, App
C 34	26.6	27.7	3608	9	US-10-175-737-433	Sequence 433, App
C 35	26.6	27.7	3608	9	US-10-175-737-433	Sequence 433, App
C 36	26.6	27.7	3608	9	US-10-173-706-433	Sequence 433, App
C 37	26.6	27.7	3608	9	US-10-175-738-433	Sequence 433, App
C 38	26.6	27.7	3608	9	US-10-175-752-433	Sequence 433, App
C 39	26.6	27.7	3608	9	US-10-176-482-433	Sequence 433, App
C 40	26.6	27.7	3608	9	US-10-176-757-433	Sequence 433, App
C 41	26.6	27.7	3608	9	US-10-176-913-433	Sequence 433, App
C 42	26.6	27.7	3608	9	US-10-180-552-433	Sequence 433, App
C 43	26.6	27.7	3608	9	US-10-180-557-433	Sequence 433, App
C 44	26.6	27.7	3608	9	US-10-173-700-433	Sequence 433, App
C 45	26.6	27.7	3608	9	US-10-174-572-433	Sequence 433, App

#### ALIGNMENTS

RESULT 1  
US-08-979-847-126  
Sequence 126, Application US/08979847  
Publication No. US20030039664A1  
GENERAL INFORMATION:  
APPLICANT: PERRON, HERVE  
APPLICANT: BESEME, FREDERIC  
APPLICANT: BRDIN, FREDERIC  
APPLICANT: PARAINOS-BACCALA, GLAUCIA  
APPLICANT: KOMURIAN-PRADEL, FLORENCE  
APPLICANT: JOLIVET-REYNAUD, COLETTE  
APPLICANT: MANDRAND, BERNARD  
APPLICANT: GARSON, JEREMY  
APPLICANT: TUBE, PHILIP  
TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS  
TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHY  
NUMBER OF SEQUENCES: 210  
CORRESPONDENCE ADDRESS:  
ADDRESSER: OLIEF & BERRIDGE, PIC  
STREET: P.O. BOX 19928  
CITY: ALEXANDRIA  
STATE: VA  
COUNTRY: USA  
ZIP: 22320  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/979,847  
FILING DATE: 26-NOV-1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: BERRIDGE, WILLIAM P.  
REGISTRATION NUMBER: 30,024  
REFERENCE/DOCKET NUMBER: WPP 39046A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703-836-6400  
TELEFAX: 703-836-2787  
INFORMATION FOR SEQ ID NO: 126:

SEQUENCE CHARACTERISTICS:  
LENGTH: 764 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
US-08-979-847-126

Query Match 70.2%; Score 67.4; DB 7; Length 764;  
Best Local Similarity 95.3%; Pred. No. 1.4e-13;  
Matches 81; Conservative 0; Mismatches 1; Indels 3; Gaps 1;

QY 12 CTCCTGAGGAGATATAATTAACACCATCTTACAGCTAGACCTCTTTGTAGAAA-- 69  
|||||  
Db 549 CTCCTGAGGAGATATAATTAACACCATCTTACAGCTAGACCTCTTTGTAGAAAAG 608  
QY 70 -AGGCAATGAGTGAAGTGCACATA 93  
|||||  
Db 609 GAGGCAATGAGTGAAGTGCACATA 633

## RESULT 2

US-08-979-847-127  
Sequence 127, Application US/08979847  
Publication No. US20030039664A1  
GENERAL INFORMATION:  
APPLICANT: PERRON, HERVE  
APPLICANT: BESEME, FREDERIC  
APPLICANT: BEDIN, FREDERIC  
APPLICANT: PARANHOS-BACCALA, GLAUCIA  
APPLICANT: KOMURIAN-PRADEL, FLORENCE  
APPLICANT: JOLIVET-REYNAUD, COLETTE  
APPLICANT: MANDRAND, BERNARD  
APPLICANT: GARSON, JEREMY  
APPLICANT: TUKE, PHILIP  
TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS  
TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACT  
NUMBER OF SEQUENCES: 210  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: OLIFF & BERRIDGE, PLC  
STREET: P.O. BOX 19928  
CITY: ALEXANDRIA  
STATE: VA  
COUNTRY: USA  
ZIP: 22320  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/979,847  
FILING DATE: 26-NOV-1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: BERRIDGE, WILLIAM P.  
REGISTRATION NUMBER: 30,024  
REFERENCE/DOCKET NUMBER: WPB 39046A  
TELEPHONE: 703-836-6400  
TELEFAX: 703-836-2787  
INFORMATION FOR SEQ ID NO: 127:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 800 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
US-08-979-847-127

Query Match 70.2%; Score 67.4; DB 7; Length 800;  
Best Local Similarity 95.3%; Pred. No. 1.5e-13;

Matches 81; Conservative 0; Mismatches 1; Indels 3; Gaps 1;

QY 12 CTCCTGAGGAGATATAATTAACACCATCTTACAGCTAGACCTCTTTGTAGAAA-- 69  
|||||  
Db 551 CTCCTGAGGAGATATAATTAACACCATCTTACAGCTAGACCTCTTTGTAGAAAAG 610  
QY 70 -AGGCAATGAGTGAAGTGCACATA 93  
|||||  
Db 611 GAGGCAATGAGTGAAGTGCACATA 635

## RESULT 3

US-08-979-847-117  
Sequence 117, Application US/08979847  
Publication No. US20030039664A1  
GENERAL INFORMATION:  
APPLICANT: PERRON, HERVE  
APPLICANT: BESEME, FREDERIC  
APPLICANT: BEDIN, FREDERIC  
APPLICANT: PARANHOS-BACCALA, GLAUCIA  
APPLICANT: KOMURIAN-PRADEL, FLORENCE  
APPLICANT: JOLIVET-REYNAUD, COLETTE  
APPLICANT: MANDRAND, BERNARD  
APPLICANT: GARSON, JEREMY  
APPLICANT: TUKE, PHILIP  
TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS  
TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYL  
NUMBER OF SEQUENCES: 210  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: OLIFF & BERRIDGE, PLC  
STREET: P.O. BOX 19928  
CITY: ALEXANDRIA  
STATE: VA  
COUNTRY: USA  
ZIP: 22320  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/979,847  
FILING DATE: 26-NOV-1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: BERRIDGE, WILLIAM P.  
REGISTRATION NUMBER: 30,024  
REFERENCE/DOCKET NUMBER: WPB 39046A  
TELEPHONE: 703-836-6400  
TELEFAX: 703-836-2787  
INFORMATION FOR SEQ ID NO: 117:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1511 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
US-08-979-847-117

Query Match 67.3%; Score 64.6; DB 7; Length 1511;  
Best Local Similarity 92.0%; Pred. No. 1.7e-12;  
Matches 80; Conservative 0; Mismatches 4; Indels 3; Gaps 1;

QY 13 TCCTGAGGAGATATAATTAACACCATCTTACAGCTAGACCTCTTTGTAGAAA---A 69  
|||||  
Db 124 TCCTGAGGAGATATAATTAACACCATCTTACAGCTAGACCTCTTTGTAGAAAAG 183  
QY 70 AGGCAATGAGTGAAGTGCACATA 96  
|||||  
Db 184 GAGGCAATGAGTGAAGTGCACATA 210



## RESULT 4

US-10-125-237-2  
: Sequence 2, Application US/10125237  
: Publication No. US20030022329A1  
: GENERAL INFORMATION:  
: APPLICANT: Tang, Y. Tom  
: APPLICANT: Liu, Chenghua  
: APPLICANT: Zhou, Ping  
: APPLICANT: Asundi, Vinod  
: APPLICANT: Ren, Feiyan  
: APPLICANT: Zhao, Qing A.  
: APPLICANT: Xue, Aidong J.  
: APPLICANT: Zhang, Jie  
: APPLICANT: Wehrman, Tom  
: APPLICANT: Wang, Jian-Rui  
: TITLE OF INVENTION: No. US20030022329A1el Nucleic Acids and  
: FILE REFERENCE: 791CIP2ADIV  
: CURRENT APPLICATION NUMBER: US/10/125,237  
: PRIOR FILING DATE: 2002-04-17  
: PRIOR APPLICATION NUMBER: 09/668,317  
: PRIOR FILING DATE: 2000-09-22  
: PRIOR APPLICATION NUMBER: 09/552,929  
: NUMBER OF SEQ ID NOS: 91  
: SOFTWARE: pt\_fl\_genes Version 2.0  
: SEQ ID NO 2  
: LENGTH: 1359  
: TYPE: DNA  
: ORGANISM: Homo sapiens  
: FEATURE:  
: NAME/KEY: CDS  
: LOCATION: (114)..(1280)  
US-10-125-237-2

Query Match 66.7%; Score 64; DB 9; Length 1359;  
Best Local Similarity 90.9%; Pred. No. 2.6e-12;  
Matches 80; Conservative 0; Mismatches 5; Indels 3; Gaps 1;

QY 12 CTCCTGAGGAGATATAATTATACACCATCTTACAGCTAGACCTCTTTGTAGAA--- 68  
DB 307 CTCCTAGAGAGATATAATTATACACCATCTTACAGCTAGACCTCTTTGTAGAAAGG 366  
QY 69 AAGGCAATGGAGTGAAGTGCATATGT 96  
DB 367 AAGGCAATGGAGTGAAGTGCATATGT 394

## RESULT 5

US-10-105-891-2  
: Sequence 2, Application US/10105891  
: Publication No. US20030073099A1  
: GENERAL INFORMATION:  
: APPLICANT: Tang, Y. Tom  
: APPLICANT: Liu, Chenghua  
: APPLICANT: Zhou, Ping  
: APPLICANT: Asundi, Vinod  
: APPLICANT: Ren, Feiyan  
: APPLICANT: Zhao, Qing A.  
: APPLICANT: Xue, Aidong J.  
: APPLICANT: Zhang, Jie  
: APPLICANT: Wehrman, Tom  
: APPLICANT: Wang, Jian-Rui  
: APPLICANT: Dimanac, Radoje T.  
: TITLE OF INVENTION: No. US20030073099A1el Nucleic Acids and  
: FILE REFERENCE: 791CIP2A  
: CURRENT APPLICATION NUMBER: US/10/105,891  
: PRIOR FILING DATE: 2002-03-25  
: PRIOR APPLICATION NUMBER: 09/668,317  
: PRIOR FILING DATE: 2000-09-22

PRIOR APPLICATION NUMBER: 09/552,929

PRIOR FILING DATE: 2000-04-18

NUMBER OF SEQ ID NOS: 91

SOFTWARE: pt\_fl\_genes Version 2.0

SEQ ID NO 2

LENGTH: 1359

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: CDS

LOCATION: (114)..(1280)

US-10-105-891-2

Query Match 66.7%; Score 64; DB 9; Length 1359;  
Best Local Similarity 90.9%; Pred. No. 2.6e-12;  
Matches 80; Conservative 0; Mismatches 5; Indels 3; Gaps 1;

QY 12 CTCCTGAGGAGATATAATTATACACCATCTTACAGCTAGACCTCTTTGTAGAA--- 68  
DB 307 CTCCTAGAGAGATATAATTATACACCATCTTACAGCTAGACCTCTTTGTAGAAAGG 366  
QY 69 AAGGCAATGGAGTGAAGTGCATATGT 96  
DB 367 AAGGCAATGGAGTGAAGTGCATATGT 394

## RESULT 6

US-09-864-761-12831/C  
: Sequence 12831, Application US/09864761  
: Patent No. US20020048763A1  
: GENERAL INFORMATION:  
: APPLICANT: Penn, Sharron G.  
: APPLICANT: Rank, David R.  
: APPLICANT: Hanzel, David K.  
: APPLICANT: Chen, Wensheng  
: TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL F  
: FILE REFERENCE: Aecolica-X-1  
: CURRENT APPLICATION NUMBER: US/09/864,761  
: PRIOR FILING DATE: 2001-05-23  
: PRIOR APPLICATION NUMBER: US 60/180,312  
: PRIOR FILING DATE: 2000-02-04  
: PRIOR APPLICATION NUMBER: US 60/207,456  
: PRIOR FILING DATE: 2000-05-26  
: PRIOR APPLICATION NUMBER: US 09/632,366  
: PRIOR FILING DATE: 2000-08-03  
: PRIOR APPLICATION NUMBER: GB 24263.6  
: PRIOR FILING DATE: 2000-10-04  
: PRIOR APPLICATION NUMBER: US 60/236,359  
: PRIOR FILING DATE: 2000-09-27  
: PRIOR APPLICATION NUMBER: PCT/US01/00666  
: PRIOR FILING DATE: 2001-01-30  
: PRIOR APPLICATION NUMBER: PCT/US01/00667  
: PRIOR FILING DATE: 2001-01-30  
: PRIOR APPLICATION NUMBER: PCT/US01/00664  
: PRIOR FILING DATE: 2001-01-30  
: PRIOR APPLICATION NUMBER: PCT/US01/00669  
: PRIOR FILING DATE: 2001-01-30  
: PRIOR APPLICATION NUMBER: PCT/US01/00665  
: PRIOR FILING DATE: 2001-01-30  
: PRIOR APPLICATION NUMBER: PCT/US01/00668  
: PRIOR FILING DATE: 2001-01-30  
: PRIOR APPLICATION NUMBER: PCT/US01/00663  
: PRIOR FILING DATE: 2001-01-30  
: PRIOR APPLICATION NUMBER: PCT/US01/00662  
: PRIOR FILING DATE: 2001-01-30  
: PRIOR APPLICATION NUMBER: PCT/US01/00661  
: PRIOR FILING DATE: 2001-01-30  
: PRIOR APPLICATION NUMBER: PCT/US01/00670  
: PRIOR FILING DATE: 2001-01-30  
: PRIOR APPLICATION NUMBER: US 60/234,687  
: PRIOR FILING DATE: 2000-09-21  
: PRIOR APPLICATION NUMBER: US 09/608,408

```
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 12831
; LENGTH: 570
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC011778.3
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 2.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 0.93
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.2
US-09-864-761-12831
```

```
Query Match          48.3%; Score 46.4; DB 10; Length 570;
Best Local Similarity 78.4%; Pred. No. 2.2e-06;
Matches 69; Conservative 0; Mismatches 16; Indels 3; Gaps 1;
```

```
OY 12 CTCCTGAGGAGTATTAATTAACACCATTTACAGCTAGACCTCTTTGT---AGAA 68
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 453 CACCTGAGGAGTATTAATTAACATCTCTGAGCTTACCTTTCTGTAAAGAG 394
```

```
OY 69 AAGGCAATGAGTGAAGTCCCATAGT 96
      | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 393 AAGGCAATGAGTGAATACATTAGT 366
```

```
RESULT 7
US-09-924-400-316/c
; Sequence 316, Application US/09924400
; Patent No. US20020165371A1
; GENERAL INFORMATION:
; APPLICANT: Fridakis, Tony N.
; APPLICANT: Reed, Steven G.
; APPLICANT: Smith, John M.
; APPLICANT: Misher, Lynda E.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Retter, Marc W.
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A. W.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Day, Craig H.
; APPLICANT: Li, Samuel X.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF BREAST CANCER
; FILE REFERENCE: 210121.419C12
; CURRENT APPLICATION NUMBER: US/09/924,400
; CURRENT FILING DATE: 2001-08-07
; NUMBER OF SEQ ID NOS: 340
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 316
; LENGTH: 584
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-924-400-316
```

```
Query Match          46.7%; Score 44.8; DB 9; Length 584;
Best Local Similarity 77.3%; Pred. No. 8e-06;
Matches 68; Conservative 0; Mismatches 17; Indels 3; Gaps 1;
```

```
OY 12 CTCCTGAGGAGTATTAATTAACACCATTTACAGCTAGACCTCTTTGTAGAA--- 68
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 339 CACCTGAGGAGTATTAATTAACATCTCTGAGCTTACCTTTCTGTAAAGAG 280
```

```
OY 69 AAGGCAATGAGTGAAGTCCCATAGT 96
      | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 279 AAGGCAATGAGTGAATACATTAGT 252
```

```
RESULT 8
US-09-810-936-316/c
; Sequence 316, Application US/09810936
; Patent No. US20020068285A1
; GENERAL INFORMATION:
; APPLICANT: Fridakis, Tony N.
; APPLICANT: Reed, Steven G.
; APPLICANT: Smith, John M.
; APPLICANT: Misher, Lynda E.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Retter, Marc W.
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A. W.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Day, Craig H.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TITLE OF INVENTION: THERAPY AND DIAGNOSIS OF BREAST CANCER
; FILE REFERENCE: 210121.419C11
; CURRENT APPLICATION NUMBER: US/09/810,936
; CURRENT FILING DATE: 2001-03-16
; NUMBER OF SEQ ID NOS: 334
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 316
; LENGTH: 584
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-810-936-316
```

```
Query Match          46.7%; Score 44.8; DB 10; Length 584;
Best Local Similarity 77.3%; Pred. No. 8e-06;
Matches 68; Conservative 0; Mismatches 17; Indels 3; Gaps 1;
```

```
OY 12 CTCCTGAGGAGTATTAATTAACACCATTTACAGCTAGACCTCTTTGTAGAA--- 68
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 339 CACCTGAGGAGTATTAATTAACATCTCTGAGCTTACCTTTCTGTAAAGAG 280
```

```
OY 69 AAGGCAATGAGTGAAGTCCCATAGT 96
      | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 279 AAGGCAATGAGTGAATACATTAGT 252
```

```
RESULT 9
US-09-924-400-307
; Sequence 307, Application US/09924400
; Patent No. US20020165371A1
; GENERAL INFORMATION:
; APPLICANT: Fridakis, Tony N.
; APPLICANT: Reed, Steven G.
; APPLICANT: Smith, John M.
; APPLICANT: Misher, Lynda E.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Retter, Marc W.
; APPLICANT: Wang, Aijun
; APPLICANT: Skeiky, Yasir A. W.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Day, Craig H.
; APPLICANT: Li, Samuel X.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF BREAST CANCER
; FILE REFERENCE: 210121.419C12
; CURRENT APPLICATION NUMBER: US/09/924,400
; CURRENT FILING DATE: 2001-08-07
; NUMBER OF SEQ ID NOS: 340
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 307
; LENGTH: 800
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-924-400-307
```

```
Query Match          46.7%; Score 44.8; DB 9; Length 800;
Best Local Similarity 77.3%; Pred. No. 8.9e-06;
Matches 68; Conservative 0; Mismatches 17; Indels 3; Gaps 1;

QY 12 CTCCTGAGGAGATTAATTATACACCATCTTACAGCTCTTTGTAGAA--- 68
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 582 CACCTGAGGAGAGTACATACATCTATCTGACCTTGACCTTTCTGTAGAGG 641
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 69 AAGGCAATGAGTGAATGACCATTAAGT 96
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 642 AAGGCAATGAGTGAATGACCATTAAGT 669

RESULT 10
US-09-825-301-11
; Sequence 11, Application US/09825301
; Patent No. US20020009738A1
; GENERAL INFORMATION:
; APPLICANT: Houghton, Raymond L.
; APPLICANT: Dillon, David C.
; APPLICANT: Molesh, David A.
; APPLICANT: Xu, Jiangchun
; APPLICANT: Zehentner, Barbara
; APPLICANT: Persing, David H.
; TITLE OF INVENTION: METHODS, COMPOSITIONS AND KITS FOR THE DETECTION
; FILE REFERENCE: 210121.513
; CURRENT APPLICATION NUMBER: US/09/825.301
; NUMBER OF SEQ ID NOS: 77
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 11
; LENGTH: 800
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-825-301-11

Query Match          46.7%; Score 44.8; DB 10; Length 800;
Best Local Similarity 77.3%; Pred. No. 8.9e-06;
Matches 68; Conservative 0; Mismatches 17; Indels 3; Gaps 1;

QY 12 CTCCTGAGGAGATTAATTATACACCATCTTACAGCTCTTTGTAGAA--- 68
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 582 CACCTGAGGAGAGTACATACATCTATCTGACCTTGACCTTTCTGTAGAGG 641
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 69 AAGGCAATGAGTGAATGACCATTAAGT 96
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 642 AAGGCAATGAGTGAATGACCATTAAGT 669

RESULT 11
US-09-810-936-307
; Sequence 307, Application US/09810936
; Patent No. US2002006285A1
; GENERAL INFORMATION:
; APPLICANT: Frudakis, Tony N.
; APPLICANT: Reed, Steven G.
; APPLICANT: Smith, John M.
; APPLICANT: Misher, Linda E.
; APPLICANT: Dillon, David C.
; APPLICANT: Reltter, Marc W.
; APPLICANT: Wang, Aljun
; APPLICANT: Skelky, Yasir A.W.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Day, Craig H.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; FILE REFERENCE: 210121.419C11
; CURRENT APPLICATION NUMBER: US/09/810.936
; NUMBER OF SEQ ID NOS: 334
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 307
```

```
; LENGTH: 800
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-810-936-307

Query Match          46.7%; Score 44.8; DB 10; Length 800;
Best Local Similarity 77.3%; Pred. No. 8.9e-06;
Matches 68; Conservative 0; Mismatches 17; Indels 3; Gaps 1;

QY 12 CTCCTGAGGAGATTAATTATACACCATCTTACAGCTCTTTGTAGAA--- 68
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 582 CACCTGAGGAGAGTACATACATCTATCTGACCTTGACCTTTCTGTAGAGG 641
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 69 AAGGCAATGAGTGAATGACCATTAAGT 96
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 642 AAGGCAATGAGTGAATGACCATTAAGT 669

RESULT 12
US-09-429-755-307
; Sequence 307, Application US/09429755A
; Patent No. US20020111467A1
; GENERAL INFORMATION:
; APPLICANT: Frudakis, Tony N.
; APPLICANT: Smith, John M.
; APPLICANT: Reed, Steven G.
; APPLICANT: Misher, Lynda
; APPLICANT: Reltter, Marc W.
; APPLICANT: Dillon, David C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; FILE REFERENCE: 210121.419C6
; CURRENT APPLICATION NUMBER: US/09/429.755A
; NUMBER OF SEQ ID NOS: 315
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 307
; LENGTH: 800
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-429-755-307

Query Match          46.7%; Score 44.8; DB 10; Length 800;
Best Local Similarity 77.3%; Pred. No. 8.9e-06;
Matches 68; Conservative 0; Mismatches 17; Indels 3; Gaps 1;

QY 12 CTCCTGAGGAGATTAATTATACACCATCTTACAGCTCTTTGTAGAA--- 68
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 582 CACCTGAGGAGAGTACATACATCTATCTGACCTTGACCTTTCTGTAGAGG 641
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 69 AAGGCAATGAGTGAATGACCATTAAGT 96
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 642 AAGGCAATGAGTGAATGACCATTAAGT 669

RESULT 13
US-09-924-400-317
; Sequence 317, Application US/09924400
; Patent No. US20020165371A1
; GENERAL INFORMATION:
; APPLICANT: Frudakis, Tony N.
; APPLICANT: Reed, Steven G.
; APPLICANT: Smith, John M.
; APPLICANT: Misher, Lynda E.
; APPLICANT: Dillon, David C.
; APPLICANT: Reltter, Marc W.
; APPLICANT: Wang, Aljun
; APPLICANT: Skelky, Yasir A.W.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Day, Craig H.
; APPLICANT: Li, Samuel X.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
```

```

; TITLE OF INVENTION: AND DIAGNOSIS OF BREAST CANCER
; FILE REFERENCE: 210121.419C12
; CURRENT APPLICATION NUMBER: US/09/924,400
; CURRENT FILING DATE: 2001-08-07
; NUMBER OF SEQ ID NOS: 340
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 317
; LENGTH: 829
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-924-400-317

Query Match          46.7%; Score 44.8; DB 9; Length 829;
Best Local Similarity 77.3%; Pred. No. 9e-06;
Matches 68; Conservative 0; Mismatches 17; Indels 3; Gaps 1;

OY 12 CTCCTGAGGAAATGATTAATATACACCATCTTACAGCTAGACCTTTGTAGAA--- 68
DB 582 CACCTGAGGAAATGATTAATATACACCATCTTACAGCTAGACCTTTGTAGAGGG 641
OY 69 AAGGCAATGAGTGAAGTGCCTAAGT 96
DB 642 AAGGCAATGAGTGAAGTGCCTAAGT 669

RESULT 14
US-09-810-936-317
; Sequence 317, Application US/09810936
; Patent No. US20020068285A1
; GENERAL INFORMATION:
; APPLICANT: Frudakis, Tony N.
; APPLICANT: Reed, Steven G.
; APPLICANT: Smith, John M.
; APPLICANT: Mishner, Linda E.
; APPLICANT: Dillon, Davin C.
; APPLICANT: Retter, Marc W.
; APPLICANT: Wang, Aijun
; APPLICANT: Skelky, Yasir A.W.
; APPLICANT: Harlocker, Susan L.
; APPLICANT: Day, Craig H.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE
; TITLE OF INVENTION: THERAPY AND DIAGNOSIS OF BREAST CANCER
; FILE REFERENCE: 210121.419C11
; CURRENT APPLICATION NUMBER: US/09/810,936
; CURRENT FILING DATE: 2001-03-16
; NUMBER OF SEQ ID NOS: 334
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 317
; LENGTH: 829
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-810-936-317

Query Match          46.7%; Score 44.8; DB 10; Length 829;
Best Local Similarity 77.3%; Pred. No. 9e-06;
Matches 68; Conservative 0; Mismatches 17; Indels 3; Gaps 1;

OY 12 CTCCTGAGGAAATGATTAATATACACCATCTTACAGCTAGACCTTTGTAGAA--- 68
DB 582 CACCTGAGGAAATGATTAATATACACCATCTTACAGCTAGACCTTTGTAGAGGG 641
OY 69 AAGGCAATGAGTGAAGTGCCTAAGT 96
DB 642 AAGGCAATGAGTGAAGTGCCTAAGT 669

RESULT 15
US-09-783-590-4203/c
; Sequence 4203, Application US/09783590
; Patent No. US20020110850A1
; GENERAL INFORMATION:
; APPLICANT: Dillion, Patrick J.
; APPLICANT: Haseltine, William A.
```

```

; APPLICANT: Li, Haodong
; APPLICANT: Rosen, Craig A.
; APPLICANT: Ruben, Steven M.
; TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2
; FILE REFERENCE: PO-16.2c1
; CURRENT APPLICATION NUMBER: US/09/783,590
; CURRENT FILING DATE: 2000-02-15
; PRIOR APPLICATION NUMBER: 08/420,856
; PRIOR FILING DATE: 1995-04-12
; PRIOR APPLICATION NUMBER: 08/346,731
; PRIOR FILING DATE: 1994-11-21
; NUMBER OF SEQ ID NOS: 12485
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4203
; LENGTH: 502
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (20)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (121)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (140)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (144)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (168)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (192)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (330)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (355)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (371)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (372)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (374)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (387)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (395)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (409)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (412)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (413)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (438)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
; LOCATION: (439)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: misc feature
```

```
LOCATION: (449)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (452)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (460)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (462)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (463)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (471)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (472)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (474)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (477)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (480)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (481)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (482)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (484)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (489)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (490)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (491)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (494)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (495)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: misc feature
LOCATION: (502)
OTHER INFORMATION: n equals a,t,g, or c
US-09-783-590-4203
```

```
Query Match 44.2%; Score 42.4; DB 10; Length 502;
Best Local Similarity 81.7%; Pred. No. 5, 1e-05;
Matches 49; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY 12 CTCCTGAGGAGTAATTAATTAACACCACTTACAGCTAGACCTTTTGTAGAAAG 71
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 83 CACCTGAGGAGTAATTAACCACTTACACCTGCACTAGATCTGTTTGTAAACAAG 24
```

Search completed: May 2, 2003, 14:43:14  
Job time : 66.8897 secs



GenCore version 5.1.5  
Copyright (c) 1993 - 2003 CompuGen Ltd.

## OM nucleic - nucleic search, using sw model

Run on: May 2, 2003, 11:21:20 ; Search time 406.862 Seconds

(Without alignments)  
6795.341 Million cell updates/sec

Title: US-09-719-554-3\_COPY\_6956\_7050

Perfect score: 95  
Sequence: 1 ccgcctcctacgcacaaagctc.....accacacagcccaacctca 95Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

## Database :

```
GenEmbl:
1:  gb_ba:*
2:  gb_htg:*
3:  gb_in:*
4:  gb_om:*
5:  gb_ov:*
6:  gb_pat:*
7:  gb_ph:*
8:  gb_pl:*
9:  gb_pr:*
10: gb_ro:*
11:  gb_sts:*
12:  gb_sy:*
13:  gb_un:*
14:  gb_vi:*
15:  em_ba:*
16:  em_fun:*
17:  em_hum:*
18:  em_in:*
19:  em_mu:*
20:  em_om:*
21:  em_or:*
22:  em_ov:*
23:  em_pat:*
24:  em_ph:*
25:  em_pl:*
26:  em_ro:*
27:  em_sts:*
28:  em_un:*
29:  em_vi:*
30:  em_htg_hum:*
31:  em_htg_in:*
32:  em_htg_other:*
33:  em_htg_mus:*
34:  em_htg_pla:*
35:  em_htg_rtd:*
36:  em_htg_mam:*
37:  em_htg_vtc:*
38:  em_sy:*
39:  em_htgo_hum:*
40:  em_htgo_mus:*
41:  em_htgo_other:*
```

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	95	100.0	3372	6	AX000964	AX000964 Sequence
2	95	100.0	3372	6	AX027478	AX027478 Sequence
3	95	100.0	3372	6	AF072502	AF072502 Homo sapi
4	95	100.0	10499	6	AX007980	AX007980 Sequence
5	95	100.0	56093	6	AX329572	AX329572 Sequence
6	95	100.0	56093	6	HSAC000064	AC000064 Human BAC
7	95	100.0	149194	9	AC007566	AC007566 Homo sapi
8	94	98.9	2930	6	AX355872	AX355872 Sequence
9	94	98.9	2930	9	AF208161	AF208161 Homo sapi
10	94	98.9	2946	6	AR177269	AR177269 Sequence
11	92.6	97.5	7582	6	AX000966	AX000966 Sequence
12	92.6	97.5	7582	6	AX027480	AX027480 Sequence
13	90.4	95.2	374	6	BD003152	BD003152 Secretary
14	90.4	95.2	374	6	I80071	I80071 Sequence 47
15	87.8	92.4	2599	6	AX007978	AX007978 Sequence
16	85.4	89.9	2006	6	AX000959	AX000959 Sequence
17	85.4	89.9	2006	6	AX027473	AX027473 Sequence
18	85.4	89.9	2006	9	AF072503	AF072503 Homo sapi
19	82.2	86.5	170754	2	AC012403	AC012403 Homo sapi
20	82.2	86.5	177866	2	AL354698	AL354698 Homo sapi
21	82.2	86.5	180954	9	AL160281	AL160281 Human DNA
22	79.2	83.4	134369	9	AC104825	AC104825 Homo sapi
23	79.2	83.4	176249	2	AC069497	AC069497 Homo sapi
24	77.6	81.7	70352	9	AL592310	AL592310 Human DNA
25	77.6	81.7	73435	2	AC091015	AC091015 Homo sapi
26	77.6	81.7	155021	9	HS774610	AL034410 Human DNA
27	77.6	81.7	159851	9	AC036114	AC036114 Homo sapi
28	77.6	81.7	163631	9	AC009276	AC009276 Homo sapi
29	77.6	81.7	165260	9	AC024341	AC024341 Homo sapi
30	77.6	81.7	168177	2	AC023914	AC023914 Homo sapi
31	77.6	81.7	170919	9	AC078847	AC078847 Homo sapi
32	77.6	81.7	186822	2	AC073235	AC073235 Homo sapi
33	77.6	81.7	195344	2	AC025248	AC025248 Homo sapi
34	77.6	81.7	196039	2	AC055757	AC055757 Homo sapi
35	76	80.0	1740	6	AX080038	AX080038 Sequence
36	76	80.0	46575	2	AC080036	AC080036 Homo sapi
37	76	80.0	114621	9	AC005187	AC005187 Homo sapi
38	76	80.0	137947	9	HS4514	AL023581 Human DNA
39	76	80.0	153444	2	AC040948	AC040948 Homo sapi
40	76	80.0	173788	9	AC022555	AC022555 Homo sapi
41	76	80.0	182928	2	AP001161	AP001161 Homo sapi
42	76	80.0	183680	9	AC098859	AC098859 Homo sapi
43	76	80.0	186540	9	AC090341	AC090341 Homo sapi
44	76	80.0	187837	9	AC079065	AC079065 Homo sapi
45	76	80.0	205035	9	CNS00005	AL049870 Human Chr

## ALIGNMENTS

```
RESULT 1
AX000964          3372 bp  DNA      linear  PAT 10-MAR-2000
LOCUS             Sequence 9 from Patent WO9902696.
DEFINITION
ACCESSION          AX000964
VERSION            AX000964.1  GI:7241206
KEYWORDS
SOURCE
ORGANISM            unidentified.
                    unidentified
REFERENCE           1 (bases 1 to 3372)
AUTHORS             Beseme,F. and Blond,J.
TITLE               ENDOGENETIC RETROVIRAL SEQUENCES, ASSOCIATED WITH AUTOIMMUNE
                    DISEASES OR WITH PREGNANCY DISORDERS
JOURNAL             Patent: WO 9902696-A 9 21-JAN-1999;
```

FEATURES	source	BIO MERIEUX (FR) / BESEME FREDERIC (FR)
LOCUS	AX027478	3372 bp DNA linear PAT 16-SEP-2000
DEFINITION	Sequence 28 from Patent FR2788784.	
ACCESSION	AX027478	
VERSION	AX027478.1 GI:10188442	
KEYWORDS	human.	
SOURCE	Homo sapiens	
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.	
REFERENCE	1 (bases 1 to 3372)	
AUTHORS	Mallet,F., Voisset,C. and Paranhos,B.G.	
JOURNAL	Patent: FR 2788784-A 28-28-JUL-2000;	
FEATURES	BIO MERIEUX (FR)	
source	Location/Qualifiers	
BASE COUNT	1047 a 835 c 711 g 779 t	
ORIGIN	1..3372 /organism="Homo sapiens" /db_xref="taxon:9606"	
Query Match	100.0%; Score 95; DB 6; Length 3372;	
Best Local Similarity	100.0%; Pred. No. 6.6e-23;	
Matches	95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
OY	1 CCTGCGCTTATGCGCCAAAGCTCCTTCAGGAAACAAAGAACAGGCGCATTAACCTCGAGAGA 60	
Db	1887 CCTGCGCTTATGCGCCAAAGCTCCTTCAGGAAACAAAGAACAGGCGCATTAACCTCGAGAGA 1946	
OY	61 CTGGCACTGATTTTATCCACAGCCCAAACTCA 95	
Db	1947 CTGGCACTGATTTTATCCACAGCCCAAACTCA 1981	
RESULT 3		
LOCUS	AF072502	3372 bp mRNA linear PRI 10-FEB-1999
DEFINITION	Homo sapiens endogenous retrovirus W sequence.	
ACCESSION	AF072502	
VERSION	AF072502.1 GI:4262286	
KEYWORDS	Homo sapiens.	
SOURCE	Homo sapiens	
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.	
REFERENCE	1 (bases 1 to 3372)	
AUTHORS	Blond,J.L., Beseme,F., Duret,L., Bouton,O., Bedin,F., Perron,H., Mandrand,B. and Mallet,F.	
JOURNAL	Molecular characterization and placental expression of HERV-W, a new human endogenous retrovirus family	
JOURNAL	J. Virol. 73 (2), 1175-1185(1999)	

MEDLINE	99090005
PUBMED	9882319
REFERENCE	2 (bases, 1 to 3372)
AUTHORS	Blond, J.-L., Beseme, F. and Mallet, F.
JOURNAL	Direct Submission
FEATURES	Submitted (19-JUN-1998) UMI03 CNRS-Biomelex, ENS Lyon, 46 allée d'Italie, Lyon, Cedex 07 69366, France
source	Location/Qualifiers
	1..3372
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/clone="cl.p15t"
	/tissue_type="Placenta"
	/clone_lib="Clontech 5'-stretch plus library, Cat number HL5014a"
misc_feature	1..3372
	/note="corresponds to pol"
repeat_region	1..3372
	/rpe_family="HERV-W"
misc_feature	2785..2786
	/note="putative splice junction"
3'UTR	3025..3351
	/note="U3-R junction undetermined"
polya_site	3350
BASE COUNT	1047 a 835 c 711 g 779 t
ORIGIN	
Query Match	100.0%; Score 95; DB 9; Length 3372;
Best Local Similarity	100.0%; Pred. No. 6.6e-23;
Matches	95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY	1 CCTGCTTATGCCCAAGCTCCTTCAGGAGAACAAAGACAGGCCATTACCTCGAGAGA 60
Db	1887 CCTGCTTATGCCCAAGCTCCTTCAGGAGAACAAAGACAGGCCATTACCTCGAGAGA 1946
OY	61 CTGGCACTGATTTTACCACCAAGCCCAACCTCA 95
Db	1947 CTGGCACTGATTTTACCACCAAGCCCAACCTCA 1981
RESULT 4	
LOCUS	AX007980 10499 bp DNA linear PAT 06-SEP-2000
DEFINITION	Sequence 3 from Patent WO967395.
ACCESSION	AX007980
VERSION	AX007980.1 GI:9995677
KEYWORDS	.
SOURCE	human.
ORGANISM	Homo sapiens
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE	1 (bases 1 to 10499)
AUTHORS	Perin, J.P., Rieger, F. and Alliel, P.M.
TITLE	Nucleic sequence and deduced protein sequence family with human endogenous retroviral motifs, and their uses
JOURNAL	Patent: WO 9667395-A 3 29-DEC-1999;
	INST NAT SANTE RECH MED (FR); PERIN JEAN PIERRE (FR); RIEGER FRANCOIS (FR); ALLIEL PATRICK M (FR)
FEATURES	Location/Qualifiers
source	1..10499
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
BASE COUNT	3048 a 2676 c 2280 g 2495 t
ORIGIN	
Query Match	100.0%; Score 95; DB 6; Length 10499;
Best Local Similarity	100.0%; Pred. No. 8.1e-23;
Matches	95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY	1 CCTGCTTATGCCCAAGCTCCTTCAGGAGAACAAAGACAGGCCATTACCTCGAGAGA 60
Db	6956 CCTGCTTATGCCCAAGCTCCTTCAGGAGAACAAAGACAGGCCATTACCTCGAGAGA 7015



Qy 61 CTGGCACTGATTTTACCACAGCCCAACCTCA 95  
 |||||||  
 Db 7016 CTGGCACTGATTTTACCACAGCCCAACCTCA 7050

RESULT 5  
 AX329572 56093 bp DNA linear PAT 09-JAN-2002  
 LOCUS Sequence 81 from Patent WO0194629.  
 DEFINITION AX329572  
 ACCESSION AX329572  
 VERSION AX329572.1 GI:18102550  
 KEYWORDS  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE  
 1 Young, P.E., Augustus, M., Carter, K.C., Ebner, R., Endress, G.,  
 Horriqan, S., Soppet, D.R. and Weaver, Z.  
 Cancer gene determination and therapeutic screening using signature  
 gene sets  
 TITLE Patent: WO 0194629-A 81 13-DEC-2001;  
 JOURNAL Avalon Pharmaceuticals (US)  
 FEATURES  
 source 1..56093  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 BASE COUNT 16164 a 12346 c 10702 g 16881 t  
 ORIGIN

Query Match 100.0%; Score 95; DB 6; Length 56093;  
 Best Local Similarity 100.0%; Pred. No. 1,1e-22;  
 Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 CCTGCTTATGCGCCAGCTCTTTCAGGAGAACAAAGAGCCATTACCTCGAGAGAGA 60  
 Db 34956 CCTGCTTATGCGCCAGCTCTTTCAGGAGAACAAAGAGCCATTACCTCGAGAGAGA 35015  
 Oy 61 CTGGCACTGATTTTACCACAGCCCAACCTCA 95  
 |||||||  
 Db 35016 CTGGCACTGATTTTACCACAGCCCAACCTCA 35050

RESULT 6  
 HSAC000064 56093 bp DNA linear PRI 13-NOV-1996  
 LOCUS Human BAC clone RG083M05 from 7q21-7q22, complete sequence.  
 DEFINITION AC000064  
 ACCESSION AC000064  
 VERSION AC000064.1 GI:1669369  
 KEYWORDS HTG.  
 SOURCE Homo sapiens.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE  
 1. (bases 1 to 56093)  
 Pauley, A.  
 TITLE The sequence of H. sapiens BAC clone RG083M05  
 JOURNAL unpublished (1996)  
 REFERENCE 2 (bases 1 to 56093)  
 WATERSTON, R.  
 TITLE Direct Submission  
 JOURNAL Submitted (13-NOV-1996)  
 COMMENT Genome Sequencing Center  
 Department of Genetics, Washington University  
 St. Louis, MO 63108, USA  
 e-mail: saplen@watson.wustl.edu

all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

SOURCE INFORMATION:  
 This clone is from the first release of the human BAC library. The library contains cloned DNA from a human male fibroblast cell line 978SK. For references see: Shizuya et al., Proc. Natl. Acad. Sci. 89:8794-8797 (1992); Kim et al., Genomics 34:213-218 (1996).  
 VECTOR: pBELO  
 Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:  
 The orientation of this clone is unknown. Actual start of this clone is at base position 1 of H.RG083M05; actual end is at 56093 of H.RG083M05

#### FEATURES

source 1..56093  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="7"  
 /map="7q21-7q22"  
 /clone="H.RG083M05"  
 /clone\_11b="C11B-978SK-B"  
 complement(838..1131)  
 /rpl\_family="ALU"  
 <1360..16971  
 /gene="WUGSC:H.RG083M05.1"  
 join(<1360..1503,4181..4370,4587..4774,6422..6556,9483..9547,11631..11773,11864..12021,13131..13296,14885..14988,16349..16546,16837..16971)  
 /gene="WUGSC:H.RG083M05.1"  
 /note="ATPase: strong similarity to peroxisome biosynthesis protein PAB1 (PID:g1172019); coded for by human CDNA C04279 (NID:g1467530)"  
 /codon\_start=1  
 /protein\_id="AAB46346.1"  
 /db\_xref="GI:1669371"  
 /translation="KRLNIOKTLVAFSEAVMOPSVLLDDLLAGLPVPEHEHSPDAORCEITCNVTKNKIDCDINKPFDLDIOHAKETGCVFARDPFLVRAIHSRLSROISITREKIVLTITDFOKALRGFLPALSLSVNLHNRDRCMRKIGLHVEYROIINDTIQUPAKVCLNKKERYPELFAPIRQRTGILLYGPGTGKTLGAVTARSRMFTISVKGPELLSKYIGASBOAVADIFIRQAAKPCILFDEFESIAPRGHDVTGTDRAVNO LITQDDGVGAGVYLAATSRPDLIDPALRPGRLDKCVCPDPDYTTISLESKTQOMLSHFVSRLEITNVLSDSLADVDLQHAVSVTDSFTGADLKALYNNOLEALHGMLSKMSRLIPDESKFNMYRILYFGSSYSESGNGTSSDLSGCLSAPESTMOPGPKKDOLFSPVPLRTASQGCCELVQEBORQDRLADISITKGYRSGSGEDSMNPGPIKTRLAISOSHLMALGHTRPISSEDMKNFAEL"

repeat\_region complement(4948..5130)  
 /rpl\_family="ALU"  
 repeat\_region complement(6581..7133)  
 /rpl\_family="11"  
 repeat\_region complement(7767..8037)  
 /rpl\_family="ALU"  
 repeat\_region complement(8186..8472)  
 /rpl\_family="ALU"  
 misc\_feature 8473..8625  
 /gene="WUGSC:H.RG083M05.1"  
 /note="match to human 3' EST H75762 (NID:g1049794), bases 287-444"  
 8841..9161  
 /gene="WUGSC:H.RG083M05.1"  
 /note="match to human 5' EST H75921 (NID:g1050050), bases 21-348"  
 9481..9547  
 /gene="WUGSC:H.RG083M05.1"  
 /note="match to human 5' EST N22627 (NID:g1130501), bases 276-343"  
 complement(12612..12907)  
 repeat\_region

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

This sequence was finished as follows unless otherwise noted:

```

misc_feature      /rpl_family="ALU"
                  /cprl70..13793
/gene="WUSC:H_RG03M05.1"
/note="match to human 5' EST H41382 (NID:917434), bases
143-266"
repeat_region    /rpl_family="ALU"
                  /cprl794..13877
/rpl_family="ALU"
13878..13906
/gene="WUSC:H_RG03M05.1"
/note="match to human 5' EST H41382 (NID:917434), bases
30-58"
repeat_region    /rpl_family="ALU"
                  /cprl907..14104
complement(14110..14137)
/rpl_family="L1"
repeat_region    complement(15618..15907)
/rpl_family="ALU"
17227..17522
/rpl_family="ALU"
18667..19235
/misc_feature     /note="match to human fetal brain 5' EST D61494
(NID:g970409), bases 1-255, and to human 3' EST R07476
(NID:g759399)"
19550..19670
/rpl_family="ALU"
21507..37303
/note="Similarity to various SS-RNA virus polypeptides;
pseudogene; region of matches and close matches to
multiple human ESTs, see R68740 (NID:984257)"
37316..37489
/note="Grail prediction, score = 80"
evidence=not_experimental
repeat_region    complement(38938..39224)
/rpl_family="ALU"
39225..39707
/misc_feature     /note="match to multiple human ESTs, see N30113
(NID:g1148633)"
39800..40085
/rpl_family="ALU"
repeat_region    complement(40247..40358)
/rpl_family="ALU"
repeat_region    complement(40632..40924)
/rpl_family="ALU"
repeat_region    complement(43283..42891)
/rpl_family="ALU"
repeat_region    complement(45474..45613)
/rpl_family="ALU"
/misc_feature     complement(45614..45737)
/note="match to human 3' EST H48898 (NID:g988738), bases
129-333"
misc_feature      complement(46107..47026)
/note="match to multiple human ESTs, see N81064
(NID:g1243765), H48897 (NID:g988737), and M78831
(NID:g273146)"
repeat_region    complement(47027..47318)
/rpl_family="ALU"
/misc_feature     complement(47365..47782)
/note="match to multiple human ESTs, see W37495
(NID:g1319089)"
47898..48115
/misc_feature     /note="match to human 5' EST H62306 (NID:g1015138), bases
93-368"
repeat_region    complement(48116..48405)
/rpl_family="ALU"
/misc_feature     complement(48406..48584)
/note="match to human 3' EST N29952 (NID:g1148472), bases
290-455, and 5' EST R12730 (NID:g765806)"
repeat_region    complement(48787..49405)
/rpl_family="ALU"
/misc_feature     complement(49406..49534)
/note="match to human 3' EST R65794 (NID:g838432), bases
309-440"
repeat_region    complement(49638..49672)

```

misc_feature	/rpt_family="A7LVU"
	complement(49674..49890)
	/note="match to human 3'
	EST N29938 (NID:g1148458), sequences are from opposite
	ends of the same clone"
gene	complement(49698..51806)
	/gene="WUSC:H_RG083M05.2"
	complement(join(49698..49888,51575..51806))
	/gene="WUSC:H_RG083M05.2"
	/note="coded for by human cDNAs W37369 (NID:g1319205),
	R65591 (NID:g838529), R65794 (NID:g838432) and R65794
	(NID:g838432)"
CDS	/codon_start=1
	/protein_id="AAB6345.1"
	/db_xref="GI:1669370"
	/translation="MFYFPCGIIICFPQGVVYOIGDVYSVIDEDQKPYAQINRGT
	IODOYEKSALTWLIPILSPROPDPASVITIEPEDLPKRMYLEFYCHAPSEPK
	SRSPPPTVPTRPEKGIVMTHTGTPATTTIKESVANLL"
exon	complement(51576..51758)
	/gene="WUSC:H_RG083M05.2"
	/note="Grail prediction, score = 86"
	/evidence="not-experimental
	complement(52052..52329)
repeat_region	/rpt_family="I1"
	55557..55843
misc_feature	/note="match to human EST M79192 (NID:g273505) base 2-289"
Query Match	100.0%; Score 95; DB 9; Length 56093;
Best Local Similarity	100.0%; Pred. No. 1.1e-22;
Matches	95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy	1 CCTGCCTTATGCCCAAGTCCTTCACGGAGAACAAAGACAGCCATTTCCTCGAGAAGA 60 
Db	34956 CCTGCCTTATGCCCAAGTCCTTCACGGAGAACAAAGACAGCCATTTCCTCGAGAAGA 35015 
Oy	61 CTGGCAACTGATTTACCACAGCCCAAACTCA 95 
Db	35016 CTGGCAACTGATTTACCACAGCCCAAACTCA 35050 
RESULT 7	
LOCUS	AC007566/c 149194 bp DNA linear PRI 01-MAR-2002
DEFINITION	Homo sapiens BAC clone CTF-1065 from 7q21-q22, complete sequence.
ACCESSION	AC007566
VERSION	AC007566.2 GI:11181861
KEYWORDS	HTG.
SOURCE	Homo sapiens.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
TITLE	1 (bases 1 to 149194)
JOURNAL	Sulston,J.E. and Waterston,R.
REFERENCE	3 (bases 1 to 149194)
AUTHORS	Waterston,R.H.
TITLE	Direct Submission
JOURNAL	Submitted (15-MAY-1999) Genome Sequencing Center, Washington
	University School of Medicine, 4444 Forest Park Parkway, St. Louis,
	MO 63108, USA
REFERENCE	4 (bases 1 to 149194)
AUTHORS	Waterston,R.
TITLE	Direct Submission
JOURNAL	Submitted (02-OCT-2000) Department of Genetics, Washington
	University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE	5 (bases 1 to 149194)

AUTHORS Waterston, R.H.  
TITLE Direct Submission  
JOURNAL Submitted (16-NOV-2000) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA  
REFERENCE 6 (bases 1 to 149194)  
AUTHORS Waterston, R.H.  
TITLE Direct Submission  
JOURNAL Submitted (03-JAN-2002) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA  
REFERENCE 7 (bases 1 to 149194)  
AUTHORS Waterston, R.H.  
TITLE Direct Submission  
JOURNAL Submitted (06-FEB-2002) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA  
REFERENCE 8 (bases 1 to 149194)  
AUTHORS Waterston, R.  
TITLE Direct Submission  
JOURNAL Submitted (01-MAR-2002) Department of Genetics, Washington University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  
COMMENT On Nov 16, 2000 this sequence version replaced g1:4835815.  
-----  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: http://genome.wustl.edu/gsc  
Contact: sapiens@wustl.wustl.edu  
Summary Statistics  
Center project name: R\_RG010605  
-----  
NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter, because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.  
  
This sequence was finished as follows unless otherwise noted:  
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality > 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.  
  
MAPPING INFORMATION:  
The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see  
http://www.nhgri.nih.gov/DIR/CTB/CHR7, send  
mailto:egreen@nhgri.nih.gov, or see http://genome.wustl.edu/gsc  
  
SOURCE INFORMATION:  
Clone CTB-10G5 is from the first release of the human BAC library CTB-978SK-B. The library contains cloned DNA from the male fibroblast cell line 978SK. See: Shizuya et al., Proc. Natl. Acad. Sci. USA 89:8794-7 (1992); U-C. Kim et al., Genomics 34:213-8 (1996). This clone is available from Research Genetics, Inc. (http://www.resgen.com).  
VECTOR: pBelorBAC11  
Selection: chloramphenicol  
  
NEIGHBORING SEQUENCE INFORMATION:  
The clone sequenced to the left is RP5-91H5, 200 base pair overlap. Actual start of this clone is at base position 195 of CTB-10G5; actual end is at base position 150532 of CTB-10G5.  
  
The clone CTB-10G5 contains the entire sequence of CTB-83M5.  
Location/Qualifiers  
1..149194

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="7"  
/map="7q21-7q22"  
/clone="CTB-10G5"  
/clone\_id="CTB-978SK-B"  
1..1634  
/rpt\_family="L1"  
2248..2388  
/note="match to EST BG752883 (NID:g14063536)"  
2248..2387  
/note="match to EST AA149693 (NID:g1720635) zn99d06.r1"  
2248..2387  
/note="match to EST AW579261 (NID:g7254310)"  
2248..2387  
/note="match to EST BG766882 (NID:g14077535)"  
2248..2387  
/note="match to EST BI160365 (NID:g14620366)"  
2248..2374  
/note="similar to Homo sapiens EST BF758865 (NID:g12106765)"  
2248..2287  
/note="match to EST AV686676 (NID:g10288539)"  
2253..2387  
/note="match to EST AU123510 (NID:g10948226)"  
2344..2387  
/note="match to EST AV686676 (NID:g10288539)"  
2696..3066  
/rpt\_family="MALR"  
3108..3392  
/rpt\_family="Alu"  
3540..3628  
/note="match to EST AA425526 (NID:g2106267) zw48b03.r1"  
3540..3628  
/note="similar to Mus musculus EST BB253526 (NID:g8946272)"  
3542..3632  
/note="match to EST AV686676 (NID:g10288539)"  
3542..3628  
/note="match to EST AA149693 (NID:g1720635) zn99d06.r1"  
3542..3628  
/note="match to EST AU123510 (NID:g10948226)"  
3542..3628  
/note="match to EST BG766882 (NID:g14077535)"  
3542..3628  
/note="match to EST BI160365 (NID:g14620366)"  
3542..3609  
/note="match to EST AW579261 (NID:g7254310)"  
3562..3628  
/note="match to EST BE272564 (NID:g9146913)"  
3598..3628  
/note="match to EST BE299708 (NID:g9183456)"  
3714..3785  
/note="match to EST BG260659 (NID:g12770475)"  
3717..3785  
/note="similar to Mus musculus EST BB994936 (NID:g10678674)"  
3719..3785  
/note="match to EST AA149693 (NID:g1720635) zn99d06.r1"  
3719..3785  
/note="match to EST AA425526 (NID:g2106267) zw48b03.r1"  
3719..3785  
/note="match to EST AU123510 (NID:g10948226)"  
3719..3785  
/note="match to EST BE272564 (NID:g9146913)"  
3719..3785  
/note="match to EST BE299708 (NID:g9183456)"  
3719..3785  
/note="match to EST BG766882 (NID:g14077535)"  
3719..3785  
/note="match to EST BI160365 (NID:g14620366)"  
3719..3785  
/note="similar to Mus musculus EST BB253526

misc\_feature (NID:98946272)"  
3860..3862  
/note="match to EST AV686676 (NID:g10288539)"  
repeat\_region 4730..4827  
/rpt\_family="L2"  
repeat\_region 4834..4925  
/rpt\_family="L1"  
misc\_feature 5715..5860  
/note="match to EST AA425526 (NID:g2106267) zw48b03.r1"  
misc\_feature 5715..5860  
/note="match to EST AU123510 (NID:g10948226)"  
misc\_feature 5715..5860  
/note="match to EST BE272564 (NID:g9146913)"  
misc\_feature 5715..5860  
/note="match to EST BE299708 (NID:g9183456)"  
misc\_feature 5715..5860  
/note="match to EST BG260659 (NID:g12770475)"  
misc\_feature 5715..5860  
/note="match to EST BG766882 (NID:g14077535)"  
misc\_feature 5715..5860  
/note="match to EST BI160365 (NID:g14620366)"  
misc\_feature 5715..5860  
/note="similar to Mus musculus EST BE994936"

Query Match 100.0%; Score 95; DB 9; Length 149194;  
Best Local Similarity 100.0%; Pred. No. 1.3e-22;  
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCGGCTTATCGCAAGCTCTCTTCAGAGAACAAAGACGGCATTACCTGGAGAGA 60  
|||||  
Db 87048 CCGGCTTATCGCAAGCTCTCTTCAGAGAACAAAGACGGCATTACCTGGAGAGA 86989  
|||||  
QY 61 CTGCAACTGATTTTACCCACAGCCCAACCTCA 95  
|||||  
Db 86988 CTGCAACTGATTTTACCCACAGCCCAACCTCA 86954  
|||||

RESULT 8  
AX355872 2930 bp DNA linear PAT 06-FEB-2002  
LOCUS AX355872  
DEFINITION Sequence 1 from Patent WO0204678.  
ACCESSION AX355872  
VERSION AX355872.1 GI:18620523  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE 1  
Keith,J.C., McCoy,J.M. and M.I.S.  
Methods and compositions for diagnosing and treating preeclampsia  
and gestational trophoblast disorders  
Patent: WO 0204678-A 17-JAN-2002;  
JOURNAL GENETICS INSTITUTE, INC. (US)  
Location/Qualifiers  
1. 2930  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
930..2546  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAD23000.1"  
/db\_xref="GI:18620524"  
/translation="MALPYHIFLFTVLPSTLTPAPPCRCMTSSSPYOEFLMROR  
GNIDAPSYSLSKGPTPTAHTHMPNCHSATLCHNANTHWGKIMNPSCPGGLGV  
TVCWTFYFTQMSDGGVODAREKHVKVISQLTRVHGTSSPYKGLDSLKHETLRT  
HRLVSLFNTLTGLHEVSAQNPNTNCMLPLNFRVYSIPVEQNNSTELINTSV  
LVGPLVSNLEITHSTNLTCVKFSNTYTNOSCIKRWVPTQIVCLPSGIFPVCSTSA  
YKLNSSSSMKFSLFVPPNTITTEODLYNYVSKPRKRPIILPFVIGAGVLAGL  
TGIGGTTSTQFYKLSOELNDMEKVAQSLVLDQQLSLAAVLAAYLQNRALDLTA  
RGCTCLFLDEECYVYNOGIVTEKYEKLRDRIQRAEELRNTGPGGLSOMPMWLP  
RSDVDIKTQPEEISAAQPLLRPNAGSS"

## FEATURES

## CDS

repeat\_region  
5'UTR  
CDS  
sig\_peptide  
mat\_peptide  
misc\_feature

BASE COUNT 842 a 800 c 571 g 717 t  
ORIGIN

Query Match 98.9%; Score 94; DB 6; Length 2930;  
Best Local Similarity 100.0%; Pred. No. 1.5e-22;  
Matches 94; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 CTGCTTATCGCAAGCTCTCTTCAGAGAACAAAGACGGCATTACCTGGAGAGAC 61  
|||||  
Db 8 CTGCTTATCGCAAGCTCTCTTCAGAGAACAAAGACGGCATTACCTGGAGAGAC 67  
|||||

QY 62 TGGCAACTGATTTTACCCACAGCCCAACCTCA 95  
|||||  
Db 68 TGGCAACTGATTTTACCCACAGCCCAACCTCA 101  
|||||

RESULT 9  
AF208161 2930 bp mRNA linear PRI 22-FEB-2000  
LOCUS AF208161  
DEFINITION Homo sapiens syncytin precursor, mRNA, complete cds.  
ACCESSION AF208161  
VERSION AF208161.1 GI:6760400  
KEYWORDS  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE 1 (bases 1 to 2930)  
Mi,S., Lee,X., Li,X., Veldman,G.M., Finnerty,H., Racie,L.,  
Lavallee,E., Tang,X.Y., Edouard,P., Howes,S., Keith,J.C. Jr. and  
McCoy,J.M.  
Syncytin is a captive retroviral envelope protein involved in human  
placental morphogenesis  
Nature 403 (6771), 785-789 (2000)  
JOURNAL 20155476  
MEDLINE 10693809  
PUBMED 2 (bases 1 to 2930)  
REFERENCE  
AUTHORS Sha,M., Lee,X., Li,X., Veldman,G.M., Finnerty,H., Racie,L.,  
Lavallee,E., Tang,X., Edouard,P., Howes,S., Keith,J.C. Jr. and  
McCoy,J.M.  
Direct Submision  
Submitted (26-NOV-1999) Genetics Institute, 87 Cambridge Park  
Drive, Cambridge, MA 02140, USA  
location/Qualifiers  
1. 2930  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="AJ172-25"  
/tissue\_type="testis"  
1. 2930  
/rpt\_family="Human endogenous retrovirus HERV-W"  
1..929  
930..2546  
/note="envelope protein"  
/product="syncytin precursor"  
/protein\_id="AAF28334.1"  
/db\_xref="GI:6760401"  
/translation="MALPYHIFLFTVLPSTLTPAPPCRCMTSSSPYOEFLMROR  
GNIDAPSYSLSKGPTPTAHTHMPNCHSATLCHNANTHWGKIMNPSCPGGLGV  
TVCWTFYFTQMSDGGVODAREKHVKVISQLTRVHGTSSPYKGLDSLKHETLRT  
HRLVSLFNTLTGLHEVSAQNPNTNCMLPLNFRVYSIPVEQNNSTELINTSV  
LVGPLVSNLEITHSTNLTCVKFSNTYTNOSCIKRWVPTQIVCLPSGIFPVCSTSA  
YKLNSSSSMKFSLFVPPNTITTEODLYNYVSKPRKRPIILPFVIGAGVLAGL  
TGIGGTTSTQFYKLSOELNDMEKVAQSLVLDQQLSLAAVLAAYLQNRALDLTA  
RGCTCLFLDEECYVYNOGIVTEKYEKLRDRIQRAEELRNTGPGGLSOMPMWLP  
RSDVDIKTQPEEISAAQPLLRPNAGSS"

repeat\_region  
5'UTR  
CDS  
sig\_peptide  
mat\_peptide  
misc\_feature

misc\_feature 2043..2120 /evidence=not\_experimental  
/note="Region: Immunosuppressive region"  
misc\_feature 2235..2357  
/note="transmembrane-region site"  
3'UTR 2547..2930  
BASE COUNT 842 a 800 c 571 g 717 t  
ORIGIN

Query Match 98.9%; Score 94; DB 9; Length 2930;  
Best Local Similarity 100.0%; Pred. No. 1.5e-22;  
Matches 94; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 CTGCTTATCGCAGCTCTCTTCAGAGAACAAAGAGCCATTACCTGAGAGAC 61  
|||||  
DB 8 CTGCTTATCGCAGCTCTCTTCAGAGAACAAAGAGCCATTACCTGAGAGAC 67  
|||||  
QY 62 TGGCAACTGATTTTACCACCAAGCCCAAACTCA 95  
|||||  
DB 68 TGGCAACTGATTTTACCACCAAGCCCAAACTCA 101  
|||||

RESULT 10  
ARI77269 2946 bp DNA linear PAT 17-DEC-2001  
LOCUS ARI77269  
DEFINITION Sequence 3 from patent US 6312921.  
ACCESSION ARI77269  
VERSION ARI77269.1 GI:17919624  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 2946)  
AUTHORS Jacoby,K., McCoy,J.M., Lavallee,E.R., Racie,L.A., Evans,C.,  
Merberg,D., M.S., and Treacy,M.  
TITLE Secreted proteins and polynucleotides encoding them  
JOURNAL Patent: US 6312921-A 3 06-NOV-2001;  
FEATURES  
source location/Qualifiers  
1..2946  
BASE COUNT 858 a 801 c 570 g 717 t  
ORIGIN

Query Match 98.9%; Score 94; DB 6; Length 2946;  
Best Local Similarity 100.0%; Pred. No. 1.5e-22;  
Matches 94; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 CTGCTTATCGCAGCTCTCTTCAGAGAACAAAGAGCCATTACCTGAGAGAC 61  
|||||  
DB 6 CTGCTTATCGCAGCTCTCTTCAGAGAACAAAGAGCCATTACCTGAGAGAC 65  
|||||  
QY 62 TGGCAACTGATTTTACCACCAAGCCCAAACTCA 95  
|||||  
DB 66 TGGCAACTGATTTTACCACCAAGCCCAAACTCA 99  
|||||

RESULT 11  
AX000966 7582 bp DNA linear PAT 10-MAR-2000  
LOCUS AX000966  
DEFINITION Sequence 11 from Patent WO9902696.  
ACCESSION AX000966  
VERSION AX000966.1 GI:7241208  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 7582)  
AUTHORS Beseme,F. and Blond,J.  
TITLE ENDOGENETIC RETROVIRAL SEQUENCES, ASSOCIATED WITH AUTOIMMUNE  
DISEASES OR WITH PREGNANCY DISORDERS  
JOURNAL Patent: WO 9902696-A 11 21-JAN-1999;  
FEATURES BIO MERIEUX (FR); BESEME FREDERIC (FR)  
location/Qualifiers

source 1..7582  
/organism="unidentified"  
/db\_xref="taxon:32644"  
BASE COUNT 2156 a 1876 c 1538 g 1796 t 216 others  
ORIGIN

Query Match 97.5%; Score 92.6; DB 6; Length 7582;  
Best Local Similarity 93.7%; Pred. No. 5.5e-22;  
Matches 89; Conservative 6; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTGCTTATCGCAGCTCTCTTCAGAGAACAAAGAGCCATTACCTGAGAGAC 60  
|||||  
DB 4657 CTGCTTATCGCAGCTCTCTTCAGAGAACAAAGAGCCATTACCTGAGAGAC 4716  
|||||  
QY 61 CTGCAACTGATTTTACCACCAAGCCCAAACTCA 95  
|||||  
DB 4717 CTGCAACTGATTTTACCACCAAGCCCAAACTCA 4751  
|||||

RESULT 12  
AX027480 7582 bp DNA linear PAT 16-SEP-2000  
LOCUS AX027480  
DEFINITION Sequence 30 from Patent FR2788784.  
ACCESSION AX027480  
VERSION AX027480.1 GI:10188444  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE 1 (bases 1 to 7582)  
AUTHORS Mallet,F., Voisset,C. and Paranhos,B.G.  
JOURNAL Patent: FR 2788784-A 30 28-JUL-2000;  
FEATURES  
source location/Qualifiers  
1..7582  
BASE COUNT 2156 a 1876 c 1538 g 1796 t 216 others  
ORIGIN

Query Match 97.5%; Score 92.6; DB 6; Length 7582;  
Best Local Similarity 93.7%; Pred. No. 5.5e-22;  
Matches 89; Conservative 6; Mismatches 0; Indels 0; Gaps 0;

QY 1 CTGCTTATCGCAGCTCTCTTCAGAGAACAAAGAGCCATTACCTGAGAGAC 60  
|||||  
DB 4657 CTGCTTATCGCAGCTCTCTTCAGAGAACAAAGAGCCATTACCTGAGAGAC 4716  
|||||  
QY 61 CTGCAACTGATTTTACCACCAAGCCCAAACTCA 95  
|||||  
DB 4717 CTGCAACTGATTTTACCACCAAGCCCAAACTCA 4751  
|||||

RESULT 13  
BD003152 374 bp DNA linear PAT 31-JAN-2002  
LOCUS BD003152  
DEFINITION Secretory proteins and polynucleotides encoding the same.  
ACCESSION BD003152  
VERSION BD003152.1 GI:18631113  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 374)  
AUTHORS McCoy,J.M., Lavallee,E.R., Racie,L.A., Merberg,D., Treacy,M.,  
Evans,C. and Spaulding,V.  
TITLE Secretory proteins and polynucleotides encoding the same  
JOURNAL Patent: JP 2001501813-A 32 13-FEB-2001;  
COMMENT GENETICS INSTITUTE INC, KENNETH JACOBS  
OS Unidentified  
PN JP 2001501813-A/32  
PD 13-FEB-2001



GenCore version 5.1.5  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 2, 2003, 11:21:20 ; Search time 151.345 Seconds

(Without alignments)  
1413.592 Million cell updates/sec

Title: US-09-719-554-3\_COPY\_6956\_7050

Perfect score: 95  
Sequence: 1 cctgcctatgcacagctc.....accacaagcccaactca 95

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

N\_Geneseq\_101002:\*

1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:\*  
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:\*  
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:\*  
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:\*  
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:\*  
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:\*  
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:\*  
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:\*  
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:\*  
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:\*  
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:\*  
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:\*  
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:\*  
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:\*  
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:\*  
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:\*  
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:\*  
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:\*  
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:\*  
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:\*  
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:\*  
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:\*  
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:\*  
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	95	100.0	3372	20	AAAX25663 Human endogenous r
2	95	100.0	3372	21	AAAX59213 Human pol gene a
3	95	100.0	3831	23	AAAS71727 DNA encoding novel
4	95	100.0	5154	23	AAAS67609 DNA encoding novel
5	95	100.0	7466	23	AAAS68626 DNA encoding novel
6	95	100.0	8279	23	AAAS76474 DNA encoding novel
7	95	100.0	8294	23	AAAS84209 DNA encoding novel
8	95	100.0	10499	21	ABN97929 Human retroviral s
9	95	100.0	56093	24	ABLC1744 Colon adenocarcino

10	94	98.9	2930	24	AAD24195 Human syncytin cdn
11	94	98.9	2945	20	AAAX77526 Human secreted pro
12	94	98.9	2946	21	AAAX59468 Human secreted pro
13	92.6	97.5	7582	20	AAAX25665 Complete human end
14	92.6	97.5	7582	21	AAAX59215 Human endogenous r
15	90.4	95.2	374	19	AAAV15135 Human adult testes
16	90.4	95.2	374	19	AAV09123 5' nucleotide sequ
17	87.8	92.4	2599	21	ABN97927 Human retroviral s
18	85.4	89.9	2006	20	AAAX25658 Human endogenous r
19	85.4	89.9	2006	21	AAAX59208 Human ovarian canc
20	81.2	85.5	337	24	ABLC80385 Human retroviral s
21	76	80.0	1740	21	ABN97973 Human endogenous r
22	74.8	78.7	2575	20	AAAX25668 Human endogenous r
23	74.4	78.3	1218	23	AAAS71717 DNA encoding novel
24	74.4	78.3	1218	23	AAAS84193 DNA encoding novel
25	74.4	78.3	2585	23	AAAS25667 Human endogenous r
26	72.8	76.6	2186	23	AAAS76465 DNA encoding novel
27	72.8	76.6	4535	23	AAAS76205 DNA encoding novel
28	71.2	74.9	758	19	AAAV43222 Multiple sclerosis
29	71.2	74.9	758	20	AAAX29705 Clone L816 from MS
30	71.2	74.9	1148	23	AAAS84200 DNA encoding novel
31	68.2	71.8	1219	23	AAAS73912 DNA encoding novel
32	63.6	66.9	1216	21	ABN97933 Human retroviral s
33	57.4	60.4	2784	21	ABN97930 Human retroviral s
34	45.2	47.6	40668	24	ABO88150 Human osteoblast d
35	40.4	42.5	747	23	AAAS91200 DNA encoding novel
36	37.2	39.2	1337	23	AAAS66625 DNA encoding novel
37	37.2	39.2	1983	23	AAAS91202 DNA encoding novel
38	35.6	37.5	5168	23	AAAS81550 Human foetal liver
39	30.8	32.4	414	22	ABAA75106 Human brain expres
40	30.8	32.4	414	22	ABA39788 Human bone marrow
41	30.8	32.4	414	22	AAK36223 Human brain expres
42	30.8	32.4	414	22	AAK49746 Human bone marrow
43	30.8	32.4	414	22	AAI26854 Probe #16787 for g
44	30.8	32.4	414	22	AAI56466 Probe #24332 used
45	30.8	32.4	414	24	ABS23239 Human genome-deriv

#### ALIGNMENTS

RESULT 1  
AAAX25663  
ID AAAX25663 standard; cDNA to mRNA; 3372 BP.  
AC AAAX25663;  
XX  
XX  
XX  
21-MAY-1999 (first entry)  
DT  
XX  
XX  
Human endogenous retrovirus w clone cl.F15T.  
DE  
XX  
XX  
Clone; human endogenous retrovirus; genome; autoimmune disease;  
KW multiple sclerosis; rheumatoid polyarthriti; insulin-dependent diabetes;  
RW disseminated lupus erythematosus; pregnancy; chromosomal marker; ss.  
KW  
XX  
XX  
Human endogenous retrovirus.  
OS  
XX  
XX  
WO9902696-A1.  
PD  
XX  
XX  
21-JAN-1999.  
PD  
XX  
XX  
06-JUL-1998; 98WO-FR01442.  
PF  
XX  
XX  
07-JUL-1997; 97FR-0008815.  
PR  
XX  
XX  
(INMR ) BIO MERIEUX.  
PA  
XX  
XX  
Beseme F, Blond JL, Bouton O, Mallet F, Mandrand B;  
PI  
XX  
XX  
WPI; 1999-120897/10.  
DR  
XX  
XX  
New nucleic acid sequences from human endogenous retrovirus-W -  
PT expressed exclusively in placenta and useful in diagnosis and

PT therapy of autoimmune disease, and abnormal or failed pregnancy  
XX  
PS Claim 1: Page 64-67; 106pp; French.  
XX  
CC This sequence represents clone cl.P15f of the human endogenous retrovirus  
CC (HERV) W genome. The nucleic acids, their fragments or peptides encoded  
CC by them are markers of autoimmune disease (e.g. multiple sclerosis,  
CC rheumatoid polyarthritis, disseminated lupus erythematosus, insulin-  
CC dependent diabetes and related pathologies) and of abnormal or  
CC unsuccessful pregnancy and can be used as chromosomal markers for  
CC susceptibility to these conditions, or proximity markers of genes  
CC associated with this susceptibility.  
XX  
SQ Sequence 3372 BP; 1047 A; 835 C; 711 G; 779 T; 0 other;  
Query Match 100.0%; Score 95; DB 20; Length 3372;  
Best Local Similarity 100.0%; Pred. No. 1.7e-22;  
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 CCTGCTTATCGCAAGCTCTTCAGAGAACAAAGAGCGCCATTACCTGGAGAGA 60  
DB 1887 CCTGCTTATCGCAAGCTCTTCAGAGAACAAAGAGCGCCATTACCTGGAGAGA 1946  
OY 61 CTGGCACTGATTATTCACCAAGCCCAACCTCA 95  
DB 1947 CTGGCACTGATTATTCACCAAGCCCAACCTCA 1981  
RESULT 2  
AAA59213  
ID AAA59213 standard; DNA; 3372 BP.  
XX  
XX AAA59213;  
AC  
XX  
DT 07-NOV-2000 (first entry)  
XX  
DE Partial pol gene and U3-R region sequences of HERV-W from human genome.  
XX  
XX Autoimmune disease; retrovirus; human endogenous retrovirus W; HERV-W;  
KW gag gene; pregnancy; multiple sclerosis; T cell proliferation; ss.  
XX  
OS Homo sapiens.  
XX  
XX WO200043521-A2.  
PN  
XX  
PD 27-JUL-2000.  
XX  
PF 21-JAN-2000; 2000WO-FR00144.  
XX  
PR 21-JAN-1999; 99PR-0000888.  
XX  
PA (INMR) BIO MERIEUX.  
XX  
PI Paranhos-Baccala G, Mallet F, Voliset C;  
XX  
DR WPI: 2000-499229/44.  
XX  
PT New nucleic acid from human endogenous retrovirus, useful e.g. for  
PT diagnosis of autoimmune disease and complications of pregnancy,  
PT contains at least part of the gag gene -  
XX  
XX  
XX Disclosure: Page 47-48; 53pp; French.  
XX  
CC The present sequence represents an endogenous retroviral nucleic acid  
CC fragment, which is associated with an autoimmune disease, and is  
CC integrated into the human genome. The fragment is originally derived  
CC from a novel retrovirus, human endogenous retrovirus W (HERV-W). The  
CC HERV-W retrovirus is associated with autoimmune disease, failure of  
CC pregnancy or disorders of pregnancy. The nucleic acid fragment, or  
CC proteins derived from it, are useful for diagnosis of autoimmune  
CC disease (specifically multiple sclerosis) and for monitoring pregnancy.  
CC The nucleic acid fragments may also be used for in situ labelling of  
CC isolated chromosomes, while the transcription product can be used to

CC study or monitor T cell proliferation in vitro.  
XX  
SQ Sequence 3372 BP; 1047 A; 835 C; 711 G; 779 T; 0 other;  
XX  
CC  
Query Match 100.0%; Score 95; DB 21; Length 3372;  
Best Local Similarity 100.0%; Pred. No. 1.7e-22;  
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 CCTGCTTATCGCAAGCTCTTCAGAGAACAAAGAGCGCCATTACCTGGAGAGA 60  
DB 1887 CCTGCTTATCGCAAGCTCTTCAGAGAACAAAGAGCGCCATTACCTGGAGAGA 1946  
OY 61 CTGGCACTGATTATTCACCAAGCCCAACCTCA 95  
DB 1947 CTGGCACTGATTATTCACCAAGCCCAACCTCA 1981  
RESULT 3  
AAS71727  
ID AAS71727 standard; CDNA; 3831 BP.  
XX  
XX AAS71727;  
AC  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE DNA encoding novel human diagnostic protein #7531.  
XX  
XX  
XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX  
OS Homo sapiens.  
XX  
XX WO200175067-A2.  
PN  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US08631.  
XX  
PR 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX  
PA (HYSE-) HISEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;  
XX  
PI  
XX  
DR WPI: 2001-639362/73.  
DR P-PSDB; ABC07540.  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
XX  
XX  
PS Claim 1: SEQ ID NO 7531; 103pp; English.  
XX  
XX  
XX The invention relates to isolated polynucleotide (I) and  
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,  
XX polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
XX and gene mapping, and in recombinant production of (II). The  
XX polynucleotides are also used in diagnostics as expressed sequence tags  
XX for identifying expressed genes. (I) is useful in gene therapy techniques  
XX to restore normal activity of (II) or to treat disease states involving  
XX (II). (II) is useful for generating antibodies against it, detecting or  
XX quantitating a polypeptide in tissue, as molecular weight markers and as  
XX a food supplement. (II) and its binding partners are useful in medical  
XX imaging of sites expressing (II). (I) and (II) are useful for treating  
XX disorders involving aberrant protein expression or biological activity.  
XX The polypeptide and polynucleotide sequences have applications in  
XX diagnostics, forensics, gene mapping, identification of mutations  
XX and to produce other types of data and products dependent on DNA and  
XX amino acid sequences. AAS64197-AAS94364 represent novel human  
XX diagnostic coding sequences of the invention.



CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 3831 BP; 1173 A; 953 C; 815 G; 890 T; 0 other;

Query Match 100.0%; Score 95; DB 23; Length 3831;  
Best Local Similarity 100.0%; Pred. No. 1.8e-22;  
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCGCCCTTATCGCCAGCCTCTTCAGAGAAACAAGAGCGCATTCCTGGAGAGA 60  
|||||  
DB 2368 CCGCCCTTATCGCCAGCCTCTTCAGAGAAACAAGAGCGCATTCCTGGAGAGA 2427  
|||||

OY 61 CTGGCAACTGATTTTACCCACAAGCCCAAACTCA 95  
|||||  
DB 2428 CTGGCAACTGATTTTACCCACAAGCCCAAACTCA 2462  
|||||

#### RESULT 4

AA567609  
ID AA567609 standard; cDNA; 5154 BP.

AC AA567609;

DT 13-FEB-2002 (first entry)

DE DNA encoding novel human diagnostic protein #3413.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;

KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

PA Drmanac RT, Liu C, Tang YT;

PI Drmanac RT, Liu C, Tang YT;

DR WPI; 2001-639362/73.

DR P-PSDB; ABG03422.

XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -

PS Claim 1; SEQ ID No 3413; 103bp; English.

XX The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and

CC amino acid sequences. AA564197-AA594564 represent novel human  
CC diagnostic coding sequences of the invention.

CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 5154 BP; 1537 A; 1321 C; 1117 G; 1179 T; 0 other;

Query Match 100.0%; Score 95; DB 23; Length 5154;  
Best Local Similarity 100.0%; Pred. No. 2e-22;  
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCGCCCTTATCGCCAGCCTCTTCAGAGAAACAAGAGCGCATTCCTGGAGAGA 60  
|||||  
DB 3691 CCGCCCTTATCGCCAGCCTCTTCAGAGAAACAAGAGCGCATTCCTGGAGAGA 3750  
|||||

OY 61 CTGGCAACTGATTTTACCCACAAGCCCAAACTCA 95  
|||||  
DB 3751 CTGGCAACTGATTTTACCCACAAGCCCAAACTCA 3785  
|||||

#### RESULT 5

AA568626  
ID AA568626 standard; cDNA; 7466 BP.

AC AA568626;

DT 13-FEB-2002 (first entry)

DE DNA encoding novel human diagnostic protein #4430.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;

KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

PF 30-MAR-2001; 2001WO-US08631.

PR 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

PA Drmanac RT, Liu C, Tang YT;

PI Drmanac RT, Liu C, Tang YT;

DR WPI; 2001-639362/73.

DR P-PSDB; ABG04439.

XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -

PS Claim 1; SEQ ID No 4430; 103bp; English.

XX The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations

CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 7466 BP; 2217 A; 1880 C; 1614 G; 1754 T; 1 other;

Query Match 100.0%; Score 95; DB 23; Length 7466;  
Best Local Similarity 100.0%; Pred. No. 2.2e-22;  
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CCTGCTTATGCGCAAGCTCTTCAGAGACAAAGAGCGCATTTACCTGGAGAGA 60  
DB 2580 CCTGCTTATGCGCAAGCTCTTCAGAGACAAAGAGCGCATTTACCTGGAGAGA 2639  
OY 61 CTGGCACTGATTTTACCCCAAGCCCAACCTCA 95  
DB 2640 CTGGCACTGATTTTACCCCAAGCCCAACCTCA 2674

RESULT 6  
AAS76474  
ID AAS76474 standard; cDNA; 8279 BP.

AC AAS76474;  
XX  
DT 13-FEB-2002 (first entry)

DE DNA encoding novel human diagnostic protein #12278.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

PE 30-MAR-2001; 2001WO-US08631.

PF 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

PA (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YT;

PI WPI; 2001-639362/73.

DR P-PSDB; ABG12287.

XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity

PS Claim 1; SEQ ID No 12278; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.

CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 8279 BP; 2474 A; 1992 C; 1770 G; 2043 T; 0 other;

Query Match 100.0%; Score 95; DB 23; Length 8279;  
Best Local Similarity 100.0%; Pred. No. 2.3e-22;  
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CCTGCTTATGCGCAAGCTCTTCAGAGACAAAGAGCGCATTTACCTGGAGAGA 60  
DB 6816 CCTGCTTATGCGCAAGCTCTTCAGAGACAAAGAGCGCATTTACCTGGAGAGA 6875  
OY 61 CTGGCACTGATTTTACCCCAAGCCCAACCTCA 95  
DB 6876 CTGGCACTGATTTTACCCCAAGCCCAACCTCA 6910

RESULT 7  
AAS84209  
ID AAS84209 standard; cDNA; 8294 BP.

AC AAS84209;  
XX  
DT 13-FEB-2002 (first entry)

DE DNA encoding novel human diagnostic protein #20013.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

XX Homo sapiens.

PN WO200175067-A2.

PD 11-OCT-2001.

PE 30-MAR-2001; 2001WO-US08631.

PF 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

PA (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YT;

PI WPI; 2001-639362/73.

DR P-PSDB; ABG20022.

XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity

PS Claim 1; SEQ ID No 20013; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical

CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations in  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. A564197-A594564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIP0  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
SQ Sequence 8294 BP; 2476 A; 1994 C; 1774 G; 2050 T; 0 other;  
Query Match 100.0%; Score 95; DB 23; Length 8294;  
Best Local Similarity 100.0%; Pred. No. 2.3e-22;  
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 CTGCGCTTATGCGCAGCTCCTTCAGAGAACAAAGACAGCCATTACCTGGAGAAGA 60  
DB 6819 CTGCGCTTATGCGCAGCTCCTTCAGAGAACAAAGACAGCCATTACCTGGAGAAGA 6878  
OY 61 CTGCGCAACTGATTTTACCACCAAGCCCAACCTCA 95  
DB 6879 CTGCGCAACTGATTTTACCACCAAGCCCAACCTCA 6913  
RESULT 8  
ABN97929  
ID ABN97929 standard; DNA; 10499 BP.  
XX  
AC ABN97929;  
XX  
DT 01-AUG-2002 (first entry)  
XX  
DE Human retroviral sequence HERV-7q.  
XX  
KW Autoimmune disease; HERV-7q; chromosome 7q; immunotherapy;  
KM multiple sclerosis; ds.  
XX  
OS Human retrovirus.  
XX  
PN WO9967395-A1.  
XX  
PD 29-DEC-1999.  
XX  
PF 23-JUN-1999; 99WO-FR01513.  
XX  
PR 23-JUN-1998; 98FR-0007920.  
XX  
PA (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.  
XX  
PI Allele PM, Perin J, Rieger F;  
XX  
DR WPI; 2000-160587/14.  
XX  
PT New nucleic acid sequences of human endogenous retrovirus, HERV-7q,  
PT used for diagnosis, treatment and prevention of autoimmune and  
PT neurological diseases  
XX  
PS Claim 3; Fig 1; 225pp; French.  
XX  
CC The present invention relates to new nucleic acid sequences of human  
CC endogenous retrovirus, HERV-7q, which is located on chromosome 7q.  
CC Regulatory elements associated with HERV-7q may alter expression of other  
CC genes (even remote genes) on the same chromosome, inducing immunological  
CC and/or neurological changes (which may be pathological or protective/  
CC curative). HERV-7q peptides can be used to improve efficiency of the  
CC immune response, e.g. in immunotherapy. HERV-7q peptides and their coding  
CC sequences can be used in immunogenic or vaccinating compositions, for  
CC protection against autoimmune diseases, particularly multiple sclerosis.  
CC The peptides may also be used (by sequence comparison) to detect/identify  
CC endogenous retroviruses that are abnormally expressed in cancer,

CC neuropathologies or other autoimmune diseases. The present sequence was  
CC used to illustrate the invention.  
XX  
SQ Sequence 10499 BP; 3048 A; 2676 C; 2280 G; 2495 T; 0 other;  
Query Match 100.0%; Score 95; DB 21; Length 10499;  
Best Local Similarity 100.0%; Pred. No. 2.5e-22;  
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 CTGCGCTTATGCGCAGCTCCTTCAGAGAACAAAGACAGCCATTACCTGGAGAAGA 60  
DB 6956 CTGCGCTTATGCGCAGCTCCTTCAGAGAACAAAGACAGCCATTACCTGGAGAAGA 7015  
OY 61 CTGCGCAACTGATTTTACCACCAAGCCCAACCTCA 95  
DB 7016 CTGCGCAACTGATTTTACCACCAAGCCCAACCTCA 7050  
RESULT 9  
ABL61744  
ID ABL61744 standard; DNA; 56093 BP.  
XX  
AC ABL61744;  
XX  
DT 15-MAY-2002 (first entry)  
XX  
DE Colon adenocarcinoma related gene sequence SEQ ID NO:81.  
XX  
KW Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;  
KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;  
KW cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;  
KW gene; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO200194629-A2.  
XX  
PD 13-DEC-2001.  
XX  
PF 30-MAY-2001; 2001WO-US10838.  
XX  
PR 05-JUN-2000; 2000US-209473P.  
PR 05-JUN-2000; 2000US-209531P.  
PR 18-SEP-2000; 2000US-231133P.  
PR 18-SEP-2000; 2000US-233617P.  
PR 20-SEP-2000; 2000US-234009P.  
PR 20-SEP-2000; 2000US-234034P.  
PR 20-SEP-2000; 2000US-234052P.  
PR 22-SEP-2000; 2000US-234509P.  
PR 22-SEP-2000; 2000US-234567P.  
PR 23-SEP-2000; 2000US-234923P.  
PR 23-SEP-2000; 2000US-234924P.  
PR 25-SEP-2000; 2000US-235077P.  
PR 25-SEP-2000; 2000US-235082P.  
PR 25-SEP-2000; 2000US-235134P.  
PR 25-SEP-2000; 2000US-235280P.  
PR 26-SEP-2000; 2000US-235637P.  
PR 26-SEP-2000; 2000US-235638P.  
PR 27-SEP-2000; 2000US-235711P.  
PR 27-SEP-2000; 2000US-235720P.  
PR 27-SEP-2000; 2000US-235840P.  
PR 27-SEP-2000; 2000US-235863P.  
PR 28-SEP-2000; 2000US-236028P.  
PR 28-SEP-2000; 2000US-236032P.  
PR 28-SEP-2000; 2000US-236033P.  
PR 28-SEP-2000; 2000US-236034P.  
PR 28-SEP-2000; 2000US-236109P.  
PR 28-SEP-2000; 2000US-236111P.  
PR 28-SEP-2000; 2000US-236842P.  
PR 29-SEP-2000; 2000US-236891P.  
PR 02-OCT-2000; 2000US-237172P.  
PR 02-OCT-2000; 2000US-237173P.  
PR 02-OCT-2000; 2000US-237278P.

PR 02-OCT-2000; 2000US-237294P.  
PR 02-OCT-2000; 2000US-237295P.  
PR 02-OCT-2000; 2000US-237316P.  
PR 03-OCT-2000; 2000US-237425P.  
PR 03-OCT-2000; 2000US-237598P.  
PR 03-OCT-2000; 2000US-237604P.  
PR 03-OCT-2000; 2000US-237606P.  
PR 03-OCT-2000; 2000US-237608P.  
PR 01-NOV-2000; 2000US-244867P.  
PR 01-NOV-2000; 2000US-245084P.  
PA (AVALON PHARM.  
PI Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;  
PI Soppet DR, Weaver Z;  
XX WPI: 2002-188264/24.  
XX  
XX  
XX Screening for anti-neoplastic agent involves exposing cells to a  
PT chemical agent to be tested for anti-neoplastic activity, and  
PT determining a change in expression of a gene of a signature gene set -  
XX  
XX Claim 1; SEQ ID 81; 44pp: English.  
XX  
XX The present invention describes a method (M1) for screening for an  
CC anti-neoplastic agent. The method involves exposing cells to a chemical  
CC agent to be tested for anti-neoplastic activity, determining a change in  
CC expression of at least one gene (I) of a signature gene set, where (I)  
CC comprises a sequence (S) selected from 8447 sequences (given in ABL61664  
CC to ABL70110), or is at least 95% identical to (S), where a change in  
CC expression is indicative of anti-neoplastic activity. (I) has cytostatic  
CC activity and can be used in gene therapy. M1 can be used for screening  
CC an anti-neoplastic agent, and can be used for producing a product which  
CC is the data collected with respect to the anti-neoplastic agent as a  
CC result of M1, and the data is sufficient to convey the chemical  
CC structure and/or properties of the agent. M1 can be used in the  
CC treatment of cancer such as colon, breast, stomach, lung, thyroid,  
CC oesophageal, ovarian, kidney, prostate or pancreatic cancer,  
CC adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,  
CC infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine  
CC carcinoma, papillary carcinoma and Wilms' tumour.  
XX  
XX Sequence 56093 BP; 16164 A; 12346 C; 10702 G; 16881 T; 0 other;  
SQ  
Query Match 100.0%; Score 95; DB 24; Length 56093;  
Best Local Similarity 100.0%; Pred. No. 4.1e-22;  
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 CCTGCTTATCGCAAGCTCTTCAGAGAACAAAGACAGGCCATTACCTGGAGAGA 60  
DB 34956 CTGCGCTTATCGCAAGCTCTTCAGAGAACAAAGACAGGCCATTACCTGGAGAGA 35015  
QY 61 CTGGCAACTGATTTTACCACCAAGCCCAACCTCA 95  
DB 35016 CTGGCAACTGATTTTACCACCAAGCCCAACCTCA 35050  
RESULT 10  
AAD24195  
ID AAD24195 standard; cDNA; 2930 BP.  
XX  
XX AAD24195;  
XX  
XX 07-MAY-2002 (first entry)  
XX  
XX Human syncytin cDNA.  
XX  
XX Human; syncytin; preclampsia; gestational trophoblast disorder;  
KW choriocarcinoma; hydatiform mole; placental site tumour; abortion;  
KW envelope gene; human endogenous defective retrovirus; HERV-W; ss.  
XX  
XX Homo sapiens.  
XX

FH Key Location/Qualifiers  
FT CDS 930..2546  
FT /\*tag= a  
FT /product= "Syncytin"  
PN WO200204678-A2.  
XX  
XX 17-JAN-2002.  
XX  
XX 09-JUL-2001; 2001WO-US21719.  
XX  
XX 07-JUL-2000; 2000US-216657P.  
XX  
XX (GENW ) GENETICS INST INC.  
XX  
XX Keith JC, McCoy JM, M1 S;  
XX  
XX WPI: 2002-171727/22.  
XX  
XX P-PSDB; AAE14540.  
XX  
XX  
XX Identifying a compound for treating a subject with or at risk of  
PT developing preclampsia, comprises determining whether the expression  
PT or activity of syncytin in the cell is modulated in the presence of a  
PT test compound -  
XX  
XX Disclosure; Page 39-42; 43pp: English.  
XX  
XX The invention relates to identifying compounds which are modulators  
CC of syncytin expression. The syncytin modulators are useful in diagnosis  
CC and treatment of preclampsia and gestational trophoblast disorders (e.g.  
CC choriocarcinoma, hydatiform mole, placental site tumour and missed/  
CC incomplete abortion). Syncytin is a human gene derived from the  
CC envelope gene of human endogenous defective retrovirus, HERV-W. The  
CC present invention is based partly on the discovery that syncytin  
CC expression is dramatically reduced in preclampsia, and is also  
CC mis-localised to the apical syncytiotrophoblast membrane. The present  
CC sequence is human syncytin cDNA.  
XX  
XX Sequence 2930 BP; 842 A; 800 C; 571 G; 717 T; 0 other;  
SQ  
Query Match 98.9%; Score 94; DB 24; Length 2930;  
Best Local Similarity 100.0%; Pred. No. 3.7e-22;  
Matches 94; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2 CTGCGCTTATCGCAAGCTCTTCAGAGAACAAAGACAGGCCATTACCTGGAGAGAC 61  
DB 8 CTGCGCTTATCGCAAGCTCTTCAGAGAACAAAGACAGGCCATTACCTGGAGAGAC 67  
QY 62 TGCGCAACTGATTTTACCACCAAGCCCAACCTCA 95  
DB 68 TGCGCAACTGATTTTACCACCAAGCCCAACCTCA 101  
RESULT 11  
AAx77526  
ID AAx77526 standard; cDNA; 2946 BP.  
XX  
XX AAx77526;  
XX  
XX 10-AUG-1999 (first entry)  
XX  
XX Human secreted protein AJ172\_2 cDNA.  
XX  
XX Secreted protein; testes; brain; blood; placenta; human; murine; thymus;  
KW bone marrow; treatment; prevention; nutrition; cytokine; immune; vaccine;  
KW cell proliferation; cell differentiation; suppressor; tumour inhibition;  
KW haematopoiesis regulator; activin; inhibitor; chemotactic; chemokine;  
KW haemostatic; thrombolytic; receptor; ligand; anti-inflammatory; tumour;  
KW cadherin; tumour invasion suppressor; gene therapy; tissue growth; ss.  
XX  
XX Homo sapiens.  
XX  
XX WO9926972-A1.  
XX  
XX PN

XX 03-JUN-1999.  
PD  
XX  
XX 17-NOV-1998; 98WO-US24614.  
PF  
XX  
XX 20-OCT-1998; 98US-0175928.  
PR  
XX 21-NOV-1997; 97US-0976110.  
PR  
XX 18-MAY-1998; 98US-0080478.  
XX  
XX (GENEY ) GENETICS INST INC.  
PA  
XX  
PI Collins-Racie LA, Evans C, Jacobs K, Lavallie ER;  
PI McCoy JM, Merberg D, Treacy M;  
XX  
XX WPI: 1999-357813/30.  
DR  
XX P-PSDB: AAY06622.  
XX  
XX New polynucleotides encoding secreted proteins  
PT  
XX  
XX Claim 13a; Page 100-101; 142pp; English.  
XX  
XX This invention describes novel human secreted proteins encoded by  
CC polynucleotides isolated from human adult testes, adult brain, adult  
CC blood or adult placenta, or murine adult bone marrow or thymus cDNA  
CC libraries. The products of the invention are predicted to have biological  
CC activities which would make them suitable for treating, preventing or  
CC ameliorating medical conditions in humans and animals, although no  
CC supporting data is given. Suggested activities include nutritional  
CC activity, cytokine and cell proliferation/differentiation activity,  
CC immune stimulating (e.g. as vaccines) or suppressing activity,  
CC haematopoiesis regulating activity, tissue growth activity,  
CC activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic  
CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory  
CC activity, cadherin/tumour invasion suppressor activity, and tumour  
CC inhibition activity. The polynucleotides are also stated to be useful  
CC for gene therapy.  
XX  
XX Sequence 2946 BP; 858 A; 801 C; 570 G; 717 T; 0 other;  
SQ

Query Match 98.9%; Score 94; DB 20; Length 2946;  
Best Local Similarity 100.0%; Pred. No. 3.7e-22;  
Matches 94; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 CTGCGTTATCGCAAGCTCTTCAGGAGAACAAAGAGCGCATTCCTGGAGAAGAC 61  
DB 6 CTGCGTTATCGCAAGCTCTTCAGGAGAACAAAGAGCGCATTCCTGGAGAAGAC 65  
OY 62 TGGCAACTGATTTTACCCACAGCCCAAACTCA 95  
DB 66 TGGCAACTGATTTTACCCACAGCCCAAACTCA 99

RESULT 12  
AAZ59468  
ID AAZ59468 standard; cDNA; 2946 BP.  
XX  
XX AAZ59468;  
AC  
XX  
XX 11-APR-2000 (first entry)  
DT  
XX  
XX Human secreted protein AJ172\_2 polynucleotide sequence.  
DE  
XX  
XX Human: secreted protein; disease diagnosis; pre-eclampsia; cancer;  
KW placental pathology; metastasis inhibition; nutritional activity;  
KW immune stimulator; haematopoiesis regulation; tissue growth;  
KW tumour inhibitor; anti-inflammatory; clone AJ172\_2; ATCC\_98115;  
KW gene therapy; ss.  
XX  
XX Homo sapiens.  
OS  
XX  
XX WO9960020-A1.  
PN  
XX  
XX 25-NOV-1999.

XX 17-MAY-1999; 99WO-US10915.  
PF  
XX  
XX 18-MAY-1998; 98US-0080478.  
PR  
XX 20-OCT-1998; 98US-0175928.  
PR  
XX  
XX (GENEY ) GENETICS INST INC.  
PA  
XX  
XX Jacobs K, McCoy JM, Lavallie ER, Collins-Racie LA, Evans C;  
PI Merberg D, Ml S, Treacy M;  
XX  
XX WPI: 2000-116311/10.  
DR  
XX P-PSDB: AAY67313.  
XX  
XX New polynucleotides encoding secreted cDNA libraries, used to develop  
PT products for the diagnosis and treatment of neoplastic disease  
XX  
XX Claim 14; Page 107-108; 149pp; English.  
XX  
XX This is the human secreted protein AJ172\_2 nucleotide sequence, obtained  
CC from a human adult testes cDNA library. The invention relates to secreted  
CC human and murine proteins. The polynucleotides and proteins are predicted  
CC to have biological activities which would make them suitable for  
CC treating, preventing or ameliorating medical conditions in humans and  
CC animals. Detection of the levels of the proteins can be used for the  
CC diagnosis of e.g. pre-eclampsia, placental pathology or cancer. Agents  
CC which modulate the expression or function of the proteins may be used for  
CC treating a neoplastic disease and inhibiting metastasis. Other suggested  
CC activities include nutritional activity (e.g. in feeds), cytokine and  
CC cell proliferation/differentiation activity, immune stimulating  
CC (e.g. as vaccines) or suppressing activity, haematopoiesis regulating  
CC activity, tissue growth activity, activin/inhibin activity,  
CC chemotactic/chemokinetic activity, haemostatic and thrombolytic activity,  
CC receptor/ligand activity, anti-inflammatory activity, cadherin/tumour  
CC invasion suppressor activity, and tumour inhibition activity. The  
CC polynucleotide sequences are also stated to be useful for gene therapy.  
XX  
XX Sequence 2946 BP; 858 A; 801 C; 570 G; 717 T; 0 other;  
SQ

Query Match 98.9%; Score 94; DB 21; Length 2946;  
Best Local Similarity 100.0%; Pred. No. 3.7e-22;  
Matches 94; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 2 CTGCGTTATCGCAAGCTCTTCAGGAGAACAAAGAGCGCATTCCTGGAGAAGAC 61  
DB 6 CTGCGTTATCGCAAGCTCTTCAGGAGAACAAAGAGCGCATTCCTGGAGAAGAC 65  
OY 62 TGGCAACTGATTTTACCCACAGCCCAAACTCA 95  
DB 66 TGGCAACTGATTTTACCCACAGCCCAAACTCA 99

RESULT 13  
AAZ55665  
ID AAZ55665 standard; cDNA to mRNA; 7582 BP.  
XX  
XX AAZ55665;  
AC  
XX  
XX 21-MAY-1999 (first entry)  
DT  
XX  
XX Complete human endogenous retrovirus W genome.  
DE  
XX  
XX Clone; human endogenous retrovirus; genome; autoimmune disease;  
KW multiple sclerosis; Rheumatoid polyarthritis; insulin-dependent diabetes;  
KW disseminated lupus erythematosus; pregnancy; chromosomal marker; ss.  
KW  
XX  
XX Human endogenous retrovirus.  
OS  
XX  
XX WO9902696-A1.  
PN  
XX  
XX 21-JAN-1999.  
PD  
XX  
XX 06-JUL-1998; 98WO-FR01442.

XX 07-JUL-1997: 97FR-0008815.  
 PR (INMR ) BIO MERIEUX.  
 PA  
 XX Beseme F, Blond JL, Bouton O, Mallet F, Mandrand B;  
 PI WPI; 1999-120897/10.  
 DR  
 XX New nucleic acid sequences from human endogenous retrovirus-W -  
 PT expressed exclusively in placenta and useful in diagnosis and  
 PT therapy of autoimmune disease, and abnormal or failed pregnancy  
 XX  
 PS Claim 1; Page 71-74; 106pp; French.  
 XX  
 CC This sequence represents the complete sequence of the human endogenous  
 CC retrovirus (HERV) W genome. The nucleic acids, their fragments or  
 CC peptides encoded by them are markers of autoimmune disease (e.g. multiple  
 CC sclerosis, rheumatoid polyarthritis, disseminated lupus erythematosus,  
 CC insulin- dependent diabetes and related pathologies) and of abnormal or  
 CC unsuccessful pregnancy and can be used as chromosomal markers for  
 CC susceptibility to these conditions, or proximity markers of genes  
 CC associated with this susceptibility.  
 XX  
 SQ Sequence 7582 BP; 2156 A; 1877 C; 1537 G; 1796 T; 2 U; 214 other;  
 Query Match 97.5%; Score 92.6; DB 20; Length 7582;  
 Best Local Similarity 93.7%; Pred. No. 1.5e-21;  
 Matches 89; Conservative 6; Mismatches 0; Indels 0; Gaps 0;  
 YY  
 CC 1 CTTCCCTTATCGCCAGCTCTCTTCAGAGAAAGAAAGCCATTACCTGGAGAGA 60  
 DB 4657 CTTCCCTTATCGCCAGCTCTCTTCAGAGAAAGAAAGCCATTACCTGGAGAGA 4716  
 YY 61 CTGGCAACTGATTTTACCACCAAGCCCAAACTCA 95  
 DB 4717 CTGGCAACTGATTTTACCACCAAGCCCAAACTCA 4751  
 XX  
 RESULT 14  
 ID AAA59215 standard; DNA: 7582 BP.  
 XX  
 AC AAA59215;  
 XX  
 DT 07-NOV-2000 (first entry)  
 DE Human endogenous retrovirus W (HERV-W) sequence.  
 XX  
 KW Autoimmune disease; retrovirus; human endogenous retrovirus W; HERV-W;  
 KW gag gene; pregnancy; multiple sclerosis; T cell proliferation; ss.  
 XX  
 OS Human endogenous retrovirus.  
 XX  
 FH Key Location/Qualifiers  
 FT LTR 1..120  
 FT /\*tag- a  
 FT /note- "R of 5' LTR"  
 FT 121..575  
 FT /\*tag- b  
 FT /note- "U5 of 5' LTR"  
 FT 579..596  
 FT primer\_bind  
 FT /\*tag- c  
 FT 581..7194  
 FT /\*tag- d  
 FT /note- "ORF1 env538"  
 FT 7039..7194  
 FT /\*tag- e  
 FT /note- "ORF2 52 AA"  
 FT 7112..7255  
 FT /\*tag- f  
 FT /note- "ORF3 48 AA"  
 FT misc\_feature 7244..7254

FT /\*tag- g  
 FT /note- "polypurine tract"  
 FT 7256..7582  
 FT LTR  
 FT /\*tag- h  
 FT /note- "U3-R of 3' LTR"  
 FT 7563..7569  
 FT /\*tag- i  
 FT 7570..7582  
 PN WO200043521-A2.  
 XX  
 PD 27-JUL-2000.  
 XX  
 PF 21-JAN-2000; 2000WO-FR00144.  
 XX  
 PR 21-JAN-1999; 99FR-0000888.  
 XX  
 PA (INMR ) BIO MERIEUX.  
 PI Paranhos-Baccala G, Mallet F, Volisset C;  
 DR WPI; 2000-499229/44.  
 XX  
 XX New nucleic acid from human endogenous retrovirus, useful e.g. for  
 PT diagnosis of autoimmune disease and complications of pregnancy,  
 PT contains at least part of the gag gene  
 XX  
 PS Disclosure; Page 49-52; 53pp; French.  
 XX  
 CC The present sequence represents an endogenous retrovirus, which is  
 CC associated with an autoimmune disease, and is integrated into the human  
 CC genome. The retrovirus is human endogenous retrovirus W (HERV-W). The  
 CC HERV-W retrovirus is associated with autoimmune disease, failure of  
 CC pregnancy or disorders of pregnancy. HERV-W nucleic acid fragments, or  
 CC proteins derived from it, are useful for diagnosis of autoimmune  
 CC disease (specifically multiple sclerosis) and for monitoring pregnancy.  
 CC The nucleic acid fragments may also be used for in situ labelling of  
 CC isolated chromosomes, while the transcription product can be used to  
 CC study or monitor T cell proliferation in vitro.  
 XX  
 SQ Sequence 7582 BP; 2156 A; 1876 C; 1538 G; 1796 T; 216 other;  
 Query Match 97.5%; Score 92.6; DB 21; Length 7582;  
 Best Local Similarity 93.7%; Pred. No. 1.5e-21;  
 Matches 89; Conservative 6; Mismatches 0; Indels 0; Gaps 0;  
 YY  
 CC 1 CTTCCCTTATCGCCAGCTCTCTTCAGAGAAAGAAAGCCATTACCTGGAGAGA 60  
 DB 4657 CTTCCCTTATCGCCAGCTCTCTTCAGAGAAAGAAAGCCATTACCTGGAGAGA 4716  
 YY 61 CTGGCAACTGATTTTACCACCAAGCCCAAACTCA 95  
 DB 4717 CTGGCAACTGATTTTACCACCAAGCCCAAACTCA 4751  
 XX  
 RESULT 15  
 ID AAV15135 standard; CDNA: 374 BP.  
 XX  
 AC AAV15135;  
 XX  
 DT 29-JUN-1998 (first entry)  
 DE Human adult testes cDNA clone A172\_2 5' cDNA portion.  
 XX  
 KW Human; adult; testes; A172\_2; secreted protein; cytokine; nutrition;  
 KW cell proliferation; differentiation; tumour; immune; ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO9804696-A1.  
 XX  
 PD 05-FEB-1998.  
 XX







GenCore version 5.1.5  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 2, 2003, 11:21:20 ; Search time 1094.14 Seconds

(without alignments)  
1406.195 Million cell updates/sec

Title: US-09-719-554-3\_COPY\_6956\_7050

Perfect score: 95

Sequence: 1 cctgcctatgcgaagctc.....accacaagcccaactca 95

Scoring table: IDENTITY\_NUC

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

EST:\*

1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estlu:\*  
4: em\_estnu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hlc:\*  
9: gb\_estl:\*  
10: gb\_estl2:\*  
11: gb\_hlc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_ylt:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_trod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	95	100.0	372	12	BF992537 IL5-GN017
2	95	100.0	379	12	BF993049 IL5-GN017
3	95	100.0	423	12	BF992535 IL5-GN017
4	95	100.0	531	12	BF012022 OV3-GN033
5	81.2	85.5	337	9	AA362784 EST72629
6	77.6	81.7	738	17	HSA408776 Homo sapi

7	70	73.7	723	13	BT72113	BT72113 603059196
8	67.8	71.4	421	12	BF761543	BF761543 CMO-CS004
9	59	62.1	352	9	AA179079	AA179079 zp1le01.r
10	55.6	58.5	508	9	AA176759	AA176759 zp1le01.s
11	42.4	44.6	660	17	AQ469362	AQ469362 CITBI-E1
12	37.2	39.2	385	17	AQ109046	AQ109046 CIT-HSP-2
13	37.2	39.2	439	17	AQ196669	AQ196669 CIT-HSP-2
14	37.2	39.2	604	14	BM713850	BM713850 UI-E-EJ0
15	35.2	37.1	489	17	B93732	B93732 CIT-HSP-216
16	34.2	36.0	188	13	BG960113	BG960113 PM3-CT064
17	33.4	35.2	540	17	AQ610340	AQ610340 HS-5093_A
18	33.2	34.9	535	17	AQ153245	AQ153245 HS-2230_B
19	32.8	34.5	674	17	A2870530	A2870530 2M0183J06
20	32.6	34.3	165	13	BG960120	BG960120 PM3-CT064
21	30	31.6	519	17	AQ753746	AQ753746 HS-2137_B
22	29.4	30.9	383	12	BQ406840	BQ406840 dact2005.
23	29.4	30.9	404	12	BF557280	BF557280 UI-R-C0-g
24	29.2	30.7	602	17	AQ234153	AQ234153 HS-2057_A
25	29.2	30.7	703	17	A2853642	A2853642 2M0156M23
26	29	30.5	676	17	AQ053323	AQ053323 CIT-HSP-2
27	29	30.5	849	17	A2898632	A2898632 RPCI-24-1
28	28.8	30.3	547	17	A2128099	A2128099 OSJNB009
29	28.8	30.3	626	10	BB636193	BB636193 BB636193
30	28.8	30.3	755	13	BQ062473	BQ062473 BQ062473
31	28.6	30.1	344	12	BE717297	BE717297 RC2-HT077
32	28.6	30.1	522	10	BE341820	BE341820 EST94654
33	28.6	30.1	597	12	BG593585	BG593585 EST492263
34	28.6	30.1	657	13	BM407602	BM407602 EST581917
35	28.6	30.1	703	12	BG590087	BG590087 EST497929
36	28.4	29.9	289	13	B1120199	B1120199 F011P78Y
37	28.4	29.9	324	17	A2628023	A2628023 1M0476A14
38	28.4	29.9	298	13	B1120140	B1120140 F011P07Y
39	28.4	29.9	353	13	B1119806	B1119806 F005P50Y
40	28.4	29.9	379	10	AM276118	AM276118 xq45h08.x
41	28.4	29.9	381	13	B1138741	B1138741 F114P91Y
42	28.4	29.9	473	13	B1127583	B1127583 G062P66Y
43	28.4	29.9	474	13	B1138181	B1138181 F100P47Y
44	28.4	29.9	490	9	A1161614	A1161614 A004P030
45	28.4	29.9	517	13	B1131573	B1131573 G122P74Y

## ALIGNMENTS

RESULT 1  
LOCUS BF992537 372 bp mRNA linear EST 23-JAN-2001  
DEFINITION IL5-GN0178-301000-203-f06 GN0178 Homo sapiens CDNA, mRNA sequence.  
ACCESSION BF992537  
VERSION BF992537.1 GI:12398860  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
1 (bases 1 to 372)  
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H., Brunstein, A., deoliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.U.  
Shotgun sequencing of the human transcriptome with ORF expressed sequence tags  
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
20202663  
Contact: Simpson A.J.U.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001

TITLE  
JOURNAL  
MEDLINE  
COMMENT

Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL:  
(http://www.ludwig.org.br/scripts/gethtml2.pl?l=IL5&t2=IL5-GN0178-301000-203-f06&t3=2000-10-30&t4=1)  
Seq primer: puc 18 forward  
High quality sequence stop: 371.

## FEATURES

source

1. 372  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="GN0178"  
/dev\_stage="Adult"  
/note="Organ: Placenta\_normal; Vector: puc18; Site\_1: SmaI  
; Site\_2: SmaI; A mini-library was made by cloning  
products derived from ORESTS PCR (U.S. Letters Patent  
Application No. 196,716 - Ludwig Institute for Cancer  
Research) profiles into the pUC 18 vector. Reverse  
transcription of tissue mRNA and cDNA amplification were  
performed under low stringency conditions."

## BASE COUNT

116 a 93 c 76 g 86 t 1 others

## ORIGIN

Query Match 100.0%; Score 95; DB 12; Length 372;  
Best Local Similarity 100.0%; Pred. No. 1.1e-20;  
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCTGCTTATCGCCAGACCTCTTCAGAGAAACAGAGCCGATTACCTGGAGAGA 60

Db 179 CTTGCTTATCGCCAGACCTCTTCAGAGAAACAGAGCCGATTACCTGGAGAGA 238

QY 61 CTGGCAACTGATTTTACCACCAAGCCCAAACTCA 95

Db 239 CTGGCAACTGATTTTACCACCAAGCCCAAACTCA 273

RESULT 2  
BF993049 379 bp mRNA linear EST 23-JAN-2001  
LOCUS IL5-GN0178-301000-204-d04 GN0178 Homo sapiens cDNA, mRNA sequence.  
DEFINITION BF993049  
ACCESSION BF993049  
VERSION BF993049.1 GI:12399372  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 379)

REFERENCE  
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,  
Nagal,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,  
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,  
Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare  
M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and  
Simpson,A.J.  
Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags  
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

TITLE  
JOURNAL MEDLINE  
COMMENT 20202663  
Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?l=IL5&t2=IL5-GN0178-  
301000-204-d04&t3=2000-10-30&t4=1)  
Seq primer: puc 18 forward  
High quality sequence stop: 323.

FEATURES

Location/Qualifiers

source

1. 379  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="GN0178"  
/dev\_stage="Adult"  
/note="Organ: Placenta\_normal; Vector: puc18; Site\_1: SmaI  
; Site\_2: SmaI; A mini-library was made by cloning  
products derived from ORESTS PCR (U.S. Letters Patent  
Application No. 196,716 - Ludwig Institute for Cancer  
Research) profiles into the pUC 18 vector. Reverse  
transcription of tissue mRNA and cDNA amplification were  
performed under low stringency conditions."

BASE COUNT 118 a 94 c 77 g 89 t 1 others

## ORIGIN

Query Match 100.0%; Score 95; DB 12; Length 379;  
Best Local Similarity 100.0%; Pred. No. 1.1e-20;  
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCTGCTTATCGCCAGACCTCTTCAGAGAAACAGAGCCGATTACCTGGAGAGA 60

Db 179 CTTGCTTATCGCCAGACCTCTTCAGAGAAACAGAGCCGATTACCTGGAGAGA 238

QY 61 CTGGCAACTGATTTTACCACCAAGCCCAAACTCA 95

Db 239 CTGGCAACTGATTTTACCACCAAGCCCAAACTCA 273

RESULT 3  
BF992535 423 bp mRNA linear EST 23-JAN-2001  
LOCUS IL5-GN0178-301000-203-e12 GN0178 Homo sapiens cDNA, mRNA sequence.  
DEFINITION BF992535  
ACCESSION BF992535  
VERSION BF992535.1 GI:12398858  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 423)

REFERENCE  
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,  
Nagal,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,  
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,  
Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare  
M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and  
Simpson,A.J.  
Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags  
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

TITLE  
JOURNAL MEDLINE  
COMMENT 20202663  
Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?l=IL5&t2=IL5-GN0178-  
301000-203-e12&t3=2000-10-30&t4=1)  
Seq primer: puc 18 forward  
High quality sequence stop: 408.

FEATURES

Location/Qualifiers

1. 423  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="GN0178"  
/dev\_stage="Adult"  
/note="Organ: Placenta\_normal; Vector: puc18; Site\_1: SmaI  
; Site\_2: SmaI; A mini-library was made by cloning  
products derived from ORESTS PCR (U.S. Letters Patent

application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT

129 a 108 c 88 g 98 t

Query Match 100.0%; Score 95; DB 12; Length 423;  
Best Local Similarity 100.0%; Pred. No. 1.1e-20;  
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCTGCTTATGCGCAAGCTCTTTCAGAGAACAAAGACAGCCATTACCTGTGAGAGA 60  
DB 179 CCTGCTTATGCGCAAGCTCTTTCAGAGAACAAAGACAGCCATTACCTGTGAGAGA 238

QY 61 CTGGCACTGATTTTACCACCAAGCCCAACCTCA 95  
DB 239 CTGGCACTGATTTTACCACCAAGCCCAACCTCA 273

RESULT 4

LOCUS BG012022 531 bp mRNA linear EST 24-JAN-2001  
DEFINITION QV3-GN0335-131200-541-d12 GN0335 Homo sapiens cDNA, mRNA sequence.

ACCESSION BG012022  
VERSION BG012022.1 GI:12460813  
KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 531)

DIAS NETO, E., GARCIA CORREA, R., VERJOVSKI-ALMEIDA, S., BRIONES, M.R., NAGAI, M.A., DA SILVA, W. Jr., ZAGO, M.A., BORDIN, S., COSTA, F.F., GOLDMAN, G.H., CARVALHO, A., MATSUKUMA, A., BAL, G.S., SIMPSON, D.H., BRUNSTEIN, A., DEOLIVEIRA, P.S., BUCHER, P., JONGENEEL, C.V., O'HARE, M.J., SOARES, F., BRENTANI, R.R., REIS, L.F., DE SOUZA, S.J. and SIMPSON, A.J.

Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

TITLE

JOURNAL

MEDLINE

COMMENT

Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?l=QV3&t2=QV3-GN0335-131200-541-d12&t3=2000-12-13&t4=1)  
Seq primer: puc 18 forward  
High quality sequence start: 18  
High quality sequence stop: 530.

FEATURES

SOURCE

1. 531

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_lib="GN0335"

/dev\_stage="Adult"

/note="Organ: Placenta,normal; Vector: puc18; Site\_1: SmaI; Site\_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent Application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT

110 a 117 c 139 g 165 t

Query Match 100.0%; Score 95; DB 12; Length 531;  
Best Local Similarity 100.0%; Pred. No. 1.2e-20;  
Matches 95; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCTGCTTATGCGCAAGCTCTTTCAGAGAACAAAGACAGCCATTACCTGTGAGAGA 60  
DB 405 CCTGCTTATGCGCAAGCTCTTTCAGAGAACAAAGACAGCCATTACCTGTGAGAGA 346

QY 61 CTGGCACTGATTTTACCACCAAGCCCAACCTCA 95  
DB 345 CTGGCACTGATTTTACCACCAAGCCCAACCTCA 311

RESULT 5

LOCUS AA362784 337 bp mRNA linear EST 21-APR-1997  
DEFINITION EST72629 Ovary II Homo sapiens cDNA 5' end, mRNA sequence.

ACCESSION AA362784  
VERSION AA362784.1 GI:2015175  
KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 337)

ADAMS, M.D., KERLAVAGE, A.R., FLEISCHMANN, R.D., FULDER, R.A., BULT, C.J., LEE, N.H., KIRKNESS, E.F., WEINSTOCK, K.G., GOCAYNE, J.D., WHITE, O., SUTTON, G., BLAKE, J.A., BRANDON, R.C., MAN-VAI, C., CLAYTON, R.A., CLINE, T.R., COTTON, M.D., EARLE-HUGHES, J., FINE, L.D., FITZGERALD, L.M., FITZHUGH, W.M., FRITCHMAN, J.L., GEORGHAN, N.S., GLODEK, A., GNEHM, C.L., HANNA, M.C., HEDBLUM, E., HINKLE, P.S. Jr., KELLEY, J.M., KELLEY, J.C., LIU, L.-I., MAMATOS, S.M., MERRICK, J.M., PELLIGRINO, S.M., MORENO-PALANQUES, R.F., McDONALD, L.A., MEYER, D.T., PELLIGRINO, S.M., PHILLIPS, C.A., RYDER, S.E., SCOTT, J.L., SAUDEK, D.M., SHIPLEY, R., SMALL, K.V., SPRIGGS, T.A., UTTERBACK, T.R., WEIDMAN, J.F., LI, Y., BEDNARIK, D.P., CAO, L., CEPEDA, M.A., COLEMAN, T.A., COLLINS, E.J., DIKKE, D., FENG, D.-F., FERRIE, A., FISCHER, C., HASTINGS, G.A., HE, W.W., HU, J.S., GREENE, J.M., GRUBER, J., HUDSON, P., KIM, A.K., KOZAK, D.L., KUNSCH, C., HUNGJUN, J., LI, H., MEISSNER, P.S., OLSEN, H., RAYMOND, L., WEI, Y.F., WING, J., XU, C., YU, G.L., RUBEN, S.M., DILLON, P.J., FANNON, M.R., ROSEN, C.A., HASSETLINE, W.A., FIELDS, C., FRASER, C.M. and VENTER, J.C.

Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence  
Nature 377 (6547 Suppl), 3-174 (1995)

TITLE

JOURNAL

MEDLINE

COMMENT

Contact: Kerlavage, AR  
Bioinformatics  
The Institute for Genomic Research  
9712 Medical Center Drive, Rockville, MD 20850 USA  
Tel: 3018699056  
Fax: 3018699423  
Email: arkerlav@tigr.org

For clone availability, additional sequence and expression information related to this EST, please check the TIGR Human Gene Index (http://www.tigr.org/tdb/hgi/hgi.html)  
Seq primer: M13 Reverse.

FEATURES

SOURCE

1. 337

/organism="Homo sapiens"

/db\_xref="ATCC (inhost):167143"

/db\_xref="taxon:9606"

/clone\_lib="Ovary II"

/dev\_stage="Female"

/sex="Female"

/note="Organ: ovary; Vector: pbluescript SK-; Site\_1: EcoRI; Site\_2: XhoI"

BASE COUNT

86 a 72 c 82 g 94 t 3 others

Query Match 85.5%; Score 81.2; DB 9; Length 337;  
Best Local Similarity 90.5%; Pred. No. 3.5e-16;  
Matches 86; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 CCGCCTTATCGCCAGCTCTCTCAGAGAACAAAGAGAGCCATTACCTGAGAGAGA 60  
|||||  
Db 184 CCGCCTTATCGCCAGCTCTCTCAGAGAACAAAGAGAGAGCCATTACCTGAGAGAGA 125  
QY 61 CTGGCAACTGATTTTACCCCAAGCCCAAACTCTCA 95  
|||||  
Db 124 CTGGCACTAGATTATTTACCCCAATGCTCA 90  
RESULT 6  
HSA408776 738 bp DNA linear GSS 02-AUG-2001  
LOCUS Homo sapiens GSS, clone BAC RPCT-11 447F21 Sp6, genomic survey  
DEFINITION  
ACCESSION AJ408776  
VERSION AJ408776.1 GI:15131033  
KEYWORDS GSS; genomic survey sequence.  
SOURCE  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
REFERENCE  
AUTHORS Beyer, K.S., Klauk, S.M., Wleemann, S. and Pousacka, A.  
TITLE Construction of a physical map of an autism susceptibility region  
JOURNAL Gene 272 (1-2), 85-91 (2001)  
MEDLINE 21363428  
PUBMED 11470513  
REFERENCE 2 (bases 1 to 738)  
AUTHORS Beyer, K.S.  
TITLE Direct Submission  
SUBMITTED (12-FEB-2001) Beyer K.S., Molecular Genome Analysis,  
Deutsches Krebsforschungszentrum, Im Neuenheimer Feld 280,  
Heidelberg, 69120, GERMANY  
FEATURES  
source location/Qualifiers  
1..738  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="7"  
/map="7q32.3-33"  
/clone="BAC RPCT-11 447F21 Sp6"  
/sex="male"  
/clone\_1lb="RPCT-11"  
BASE COUNT 241 a 174 c 160 g 163 t  
ORIGIN  
Query Match 81.7%; Score 77.6; DB 17; Length 738;  
Best Local Similarity 94.8%; Pred. No. 6.8e-15;  
Matches 91; Conservative 0; Mismatches 4; Indels 1; Gaps 1;  
QY 1 CCGCCTTATCGCCAGCTCTCTCAGAGAACAAAGAGAGCCATTACCTGAGAGAGA 60  
|||||  
Db 507 CCGCCTTATCGCCAGCTCTCTCAGAGAACAAAGAGAGAGCCATTACCTGAGAGAGA 566  
QY 61 CTGGCAACTGA-TTTACCCCAAGCCCAAACTCTCA 95  
|||||  
Db 567 CTGGCAACTAGATTATTTACCCCAATGCTCA 602  
RESULT 7  
B1772113 723 bp mRNA linear EST 25-SEP-2001  
LOCUS B1772113  
DEFINITION 603059196F1 NIH\_MGC\_122 Homo sapiens CDNA IMAGE:5208814 5',  
mRNA sequence.  
ACCESSION B1772113  
VERSION B1772113.1 GI:15763691  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
REFERENCE 1 (bases 1 to 723)

AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgabbs-remail.nih.gov](mailto:cgabbs-remail.nih.gov)  
Tissue Procurement: Life Technologies, Inc.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: L1AM1523 row: 0 column: 23  
High quality sequence stop: 722.  
FEATURES  
source location/Qualifiers  
1..723  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_1lb="NIH\_MGC\_122"  
/lab\_host="DH10B"  
/note="Organ: pooled lung and spleen; Vector: pCMV-SPORT6;  
Site\_1: Not; Site\_2: EcoRV (destroyed); RNA source  
anonymous pool of 24 week female lung, 16 week female  
spleen, and 20-22 week male spleens. Library is oligo-dT  
primed and directionally cloned (EcoRV site is destroyed  
upon cloning). Average insert size 1.4 kb, insert size  
range 1-3 kb. Library is normalized and enriched for  
full-length clones and was constructed by C. Gruber  
(Invitrogen). Research Genetics tracking code 026. Note:  
this is a NIH-MGC Library."  
BASE COUNT 229 a 185 c 141 g 168 t  
ORIGIN  
Query Match 73.7%; Score 70; DB 13; Length 723;  
Best Local Similarity 84.0%; Pred. No. 2.1e-12;  
Matches 79; Conservative 0; Mismatches 15; Indels 0; Gaps 0;  
QY 1 CCGCCTTATCGCCAGCTCTCTCAGAGAACAAAGAGAGCCATTACCTGAGAGAGA 60  
|||||  
Db 121 CCGCCTTATCGCCAGCTCTCTCAGAGAACAAAGAGAGAGCCATTACCTGAGAGAGA 180  
QY 61 CTGGCAACTGATTTTACCCCAAGCCCAAACTCTC 94  
|||||  
Db 181 CTGGCAACTAGATTATTTACCCCAATGCTCA 214  
RESULT 8  
BF761543 421 bp mRNA linear EST 12-JAN-2001  
LOCUS BF761543  
DEFINITION CM0-CS0041-051000-592-c12 CS0041 Homo sapiens CDNA, mRNA sequence.  
ACCESSION BF761543  
VERSION BF761543.1 GI:12109443  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
REFERENCE 1 (bases 1 to 421)  
AUTHORS Dias Neto, E., Garcia Correa, R., Verjowski-Almeida, S., Briones, M.R.,  
Nagai, M.A., da Silva, M. Jr., Zago, M.A., Bordin, S., Costa, F.F.,  
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D.H.,  
Brunstein, A., deoliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,  
M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and  
Simpson, A.J.  
Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags  
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
JOURNAL 20202663  
MEDLINE  
COMMENT Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,

Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?tbl=CM0&ct=CM0-CS0041-051000-592-cl2&ct=2000-10-05&ct4=1)  
Seq primer: puc 18 forward  
High quality sequence stop: 11  
Location/Qualifiers  
1. 421  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="CS0041"  
/note="Organ: colon\_est; Vector: puc18; Site\_1: SmaI; Site\_2: SmaI; A mini-library was made by cloning products derived from ORESRES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."  
BASE COUNT 122 a 111 c 86 g 102 t  
ORIGIN  
Query Match 71.4%; Score 67.8; DB 12; Length 421;  
Best Local Similarity 82.1%; Pred. No. 9.2e-12;  
Matches 78; Conservative 0; Mismatches 17; Indels 0; Gaps 0;  
QY 1 CCGCCTTATCGCAGCCTCTCAGAGAAAGAACAGCCATTACCTGGAGAGA 60  
DB 258 CCGCCTTATCGCAGCCTCTCAGAGAAAGAACAGCCATTACCGAGAGACAC 317  
QY 61 CTGGCAACTGATTTTACCAGCAAGCCCAACTCA 95  
DB 318 TGGCACTAGATTTTTACCAGCAAGCCCAACTCA 352  
RESULT 9  
LOCUS AAI179079 352 bp mRNA linear EST 31-DEC-1996  
DEFINITION zplie01.r1 Stratagene fetal retina 937202 Homo sapiens cDNA clone IMAGE:609144 5' similar to SW:POL\_AVRHE P03360 POL POLYPROTEIN ;  
ACCESSION AAI179079  
VERSION AAI179079.1 GI:1760431  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 352)  
Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B., Chisoe,S., Dietrich,N., Dubuque,T., Favello,A., Gish,W., Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N., Maris,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Mieg,J., Treviskis,E., Underwood,K., Wohldmann,P., Waterston,R., Wilson,R. and Marra,M. Generation and analysis of 280,000 human expressed sequence tags Genome Res. 6 (9), 807-828 (1996)  
9704478  
TITLE Contact: Wilson RK  
JOURNAL Washington University School of Medicine  
MEDLINE 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
COMMENT Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@wustl.wustl.edu  
This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
Seq primer: -28M13 rev2 from Amersham  
High quality sequence stop: 320.

FEATURES  
source Location/Qualifiers  
1. 352  
/organism="Homo sapiens"  
/db\_xref="GDB:4624403"  
/db\_xref="taxon:9606"  
/clone\_lib="IMAGE:609144"  
/clone\_lib="Stratagene fetal retina 937202"  
/sex="mixed"  
/lab\_host="SOLR (kanamycin resistant)"  
/note="Vector: pbluescript SK-; Site\_1: EcoRI; Site\_2: XhoI; Cloned unidirectionally. Primer: Oligo dT. Pooled retinal tissue. Average insert size: 1.0 kb; Uni-ZAP XR Vector: -5' adaptor sequence: 5' GATTGGGACGAG 3' -3' adaptor sequence: 5' CTCGAGTTTCTTTTCTTTTCTTTT 3'"  
BASE COUNT 114 a 88 c 65 g 80 t 5 others  
ORIGIN  
Query Match 62.1%; Score 59; DB 9; Length 352;  
Best Local Similarity 91.3%; Pred. No. 6.6e-09;  
Matches 84; Conservative 0; Mismatches 6; Indels 2; Gaps 2;  
QY 1 CCGCCTTATCGCAGCCTCTCAGAGAAAGAACAGGCA-TTACCTGGAGAG 59  
DB 247 CCGCCTTATCGCAGCCTCTCAGAGAAAGAACAGGCAATTACCGAGAGAG 306  
QY 60 ACTGGCACT-GATTTTACCAGAGCCCAA 90  
DB 307 ACTGGCACTGATTTTACCAGAGCCCAA 338  
RESULT 10  
LOCUS AAI16759/c 508 bp mRNA linear EST 30-DEC-1996  
DEFINITION zplie01.s1 Stratagene fetal retina 937202 Homo sapiens cDNA clone IMAGE:609144 3' similar to SW:POL\_MLYRK P31795 POL POLYPROTEIN ;  
ACCESSION AAI16759  
VERSION AAI16759.1 GI:1758064  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 508)  
Hillier,L., Lennon,G., Becker,M., Bonaldo,M.F., Chiapelli,B., Chisoe,S., Dietrich,N., Dubuque,T., Favello,A., Gish,W., Hawkins,M., Hultman,M., Kucaba,T., Lacy,M., Le,M., Le,N., Maris,E., Moore,B., Morris,M., Parsons,J., Prange,C., Rifkin,L., Rohlfing,T., Schellenberg,K., Soares,M.B., Tan,F., Thierry-Mieg,J., Treviskis,E., Underwood,K., Wohldmann,P., Waterston,R., Wilson,R. and Marra,M. Generation and analysis of 280,000 human expressed sequence tags Genome Res. 6 (9), 807-828 (1996)  
9704478  
TITLE Contact: Wilson RK  
JOURNAL Washington University School of Medicine  
MEDLINE 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
COMMENT Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@wustl.wustl.edu  
This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
Seq primer: -40M13 fwd. from Amersham  
High quality sequence stop: 270.  
FEATURES  
source Location/Qualifiers  
1. 508  
/organism="Homo sapiens"  
/db\_xref="GDB:4624403"  
/db\_xref="taxon:9606"  
/clone\_lib="IMAGE:609144"  
/clone\_lib="Stratagene fetal retina 937202"  
/sex="mixed"  
/lab\_host="SOLR (kanamycin resistant)"  
/note="Vector: pbluescript SK-; Site\_1: EcoRI; Site\_2:

Query Match 58.5%; Score 55.6; DB 9; Length 508;  
Best Local Similarity 86.7%; Pred. No. 9,7e-08;  
Matches 72; Conservative 0; Mismatches 10; Indels 1; Gaps 1;

BASE COUNT 112 a 99 c 130 g 163 t 4 others

ORIGIN

xhoI: Cloned unidirectionally. Primer: Oligo dT. Pooled  
retinal tissue. Average insert size: 1.0 kb; Uni-ZAP XR  
Vector: -5' adaptor sequence: 5' GATTCCGCGACAG 3' -3'  
adaptor sequence: 5' CTCGAGTTTCTTTTCTTTT 3'

QY 1 CCTGCTTATCGCAAGCTCTTCAGAGAA-CAAGAAGACGTCATTACCTCGAAG 59  
DB 482 CTTCGCTTATCACCACCAAGCCTTCAGAGAACCAAGAGCCATTACCGAGAG 423

QY 60 ACTGCAACTGATTATTCACCA 82  
DB 422 ACTGCAACTGATTATTCACCA 400

RESULT 11  
LOCUS A0469362/c 660 bp DNA linear GSS 23-APR-1999

DEFINITION C17B1-E1-2595M14.TR C17B1-E1 Homo sapiens genomic clone 2595M14,  
DNA sequence.

ACCESSION A0469362  
VERSION A0469362.1 GI:4653252

KEYWORDS GSS.

SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 660)  
Zhang, S., Adams, M.D., Nierman, W., Malek, J., Shizuya, H., Simon, M. and  
Venter, J.C.

TITLE Use of BAC End Sequences from Caltech Libraries for Sequence-Ready  
Map Building

JOURNAL Unpublished (1997)

COMMENT Other\_GSSs: C17B1-E1-2595M14.TF  
Contact: Shaying Zhao, William Nierman, Mark Adams  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: hbe@tigr.org

Clones are available from Research Genetics (Info@resgen.com). BAC  
end search page:  
http://www.tigr.org/tdb/humgen/bac\_end\_search/bac\_end\_search.html.  
Seq primer: M13 Reverse  
Class: BAC ends.

FEATURES  
Location/Qualifiers  
1..660  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="2595M14"  
/clone\_lib="C17B1-E1"  
/sex="male"  
/cell\_type="sperm"  
/note="Vector: pBelobAC11; Site\_1: EcoRI; Site\_2: EcoRI;  
Caltech Human BAC Library D"

BASE COUNT 158 a 137 c 146 g 219 t

ORIGIN

Query Match 44.6%; Score 42.4; DB 17; Length 660;  
Best Local Similarity 72.4%; Pred. No. 0.0022;  
Matches 55; Conservative 0; Mismatches 21; Indels 0; Gaps 0;

QY 15 AAGCTCTTCAGAGAACAGAGCCATTACCTGAGAGAGCACTGCACTGATT 74  
DB 369 AGGCTCTTCAG 310

QY 75 TACCCACAAGCCCAA 90

DB 309 TTACCCATATGCCAA 294

RESULT 12  
LOCUS A0109046/c 385 bp DNA linear GSS 29-AUG-1998

DEFINITION C17-HSP-2377A19.TR C17-HSP Homo sapiens genomic clone 2377A19, DNA  
sequence.

ACCESSION A0109046  
VERSION A0109046.1 GI:3485736

KEYWORDS GSS.

SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 385)  
Berry, K., Rounsley, S.D., Zhao, S., Bass, S., Linher, K., Golden, K.,  
Berry, K., Granger, D., Suh, E., Wible, C., Shizuya, H., Simon, M. and  
Venter, J.C.

TITLE Use of a random human BAC End Sequence Database for Sequence-Ready  
Map Building

JOURNAL Unpublished (1998)

COMMENT Other\_GSSs: C17-HSP-2377A19.TF  
Contact: Mark Adams  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: mdamas@tigr.org

Clones are available from Research Genetics (Info@resgen.com). BAC  
end search page:  
http://www.tigr.org/tdb/humgen/bac\_end\_search/bac\_end\_search.html.  
Seq primer: M13 Reverse  
Class: BAC ends.

FEATURES  
Location/Qualifiers  
1..385  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="2377A19"  
/clone\_lib="C17-HSP"  
/sex="Male"  
/cell\_type="sperm"  
/note="Vector: pBelobAC11; Site\_1: HindIII; Site\_2:  
HindIII"

BASE COUNT 100 a 77 c 88 g 120 t

ORIGIN

Query Match 39.2%; Score 37.2; DB 17; Length 385;  
Best Local Similarity 65.9%; Pred. No. 0.094;  
Matches 54; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

QY 9 ATGCCCAAGCTCTTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 68  
DB 223 ATCATAG 164

QY 69 TGATTTTACCCACAAGCCCAA 90  
DB 163 TAGACTTCACCATATGCTTA 142

RESULT 13  
LOCUS A0196669/c 439 bp DNA linear GSS 16-SEP-1998

DEFINITION C17-HSP-2381B2.TR C17-HSP Homo sapiens genomic clone 2381B2, DNA  
sequence.

ACCESSION A0196669  
VERSION A0196669.1 GI:3604031

KEYWORDS GSS.

SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

FEATURES		Source	
<p>CDNA Library preparation: Dr. M. Bento Soares, University of Iowa            CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa            DNA Sequencing by: Dr. M. Bento Soares, University of Iowa            Clone Distribution: Researchers may obtain clones from Research            Genetics (www.resgen.com).            The following repetitive elements were found in this CDNA            sequence: 2.604.&gt;HERV9#LTR/Retroviral            Seq primer: M13 Reverse.</p>			
Location/Qualifiers			
1..604			
/organism="Homo sapiens"			
/db_xref="taxon:9606"			
/clone="UI-E-EJ0-ahp-j-24-0-UI"			
/clone_1lb="UI-E-EJ0"			
/tissue_type="fetal eyes, lens, eye anterior segment, optic nerve, retina, Retina foveal and Macular, RPE and Choroid"			
/dev_stage="fetal and adult"			
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"			
/note="Organ: eye; Vector: pT7T3-Pac (Pharmacia) with a modified polylinker; Site 1: Ecor I; Site 2: Not I; UI-E-EJ0 is a substracted cDNA library constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an Ecor I adaptor, digested with Not I, and cloned directionally into pT7T3-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tags for this library are: fetal eyes, AGAATCAACA ; lens, CGATTAGCGA; eye anterior segment, AATGCCGAT; optic nerve, CCATTAGTGT; retina, CCGCG; Retina foveal and Macular, GTCC; RPE and Choroid, ACCCTA. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NBI)."			
BASE COUNT			
183 a 139 c 134 g 148 t			
ORIGIN			
Query Match			
Best Local Similarity 65.9%; Pred. No. 0.11;			
Matches 54; Conservative 0; Mismatches 28; Indels 0; Gaps 0;			
QY	9	ATGCCCAAGCTCTTTCAGGAGAACAAAGACAGCCATTACCTCGGAGAAACACTGGCAC	68
Db	316	ATCATTAAGGCCCCCTTTGGGGGAGACAAAGAAATATGCTACTATCTTGGAGAGAGCTGGCAGT	375
QY	69	TGATTTACCCCAAGCCCAAA	90
Db	376	TAGACTTACCCCATATGCTTA	397
RESULT 15			
LOCUS	B93732	489 bp	DNA linear GSS 25-JUN-1998
DEFINITION			
CIT-HSP-2165119.TR CIT-HSP Homo sapiens genomic clone 2165119, DNA			
sequence.			
ACCESSION	B93732		
VERSION	B93732.1		
KEYWORDS	GSS.		
SOURCE	human.		
ORGANISM	Homo sapiens		
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
REFERENCE	1 (bases 1 to 489)		
AUTHORS	Adams,M.D., Rounsley,S.D., Zhao,S., Field,C.E., Bass,S., Linher,K., Golden,K., Berry,K., Granger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and Verter,J.C.		
TITLE	Use of a random BAC End Sequence Database for Sequence-Ready Map Building (1998)		
JOURNAL	Unpublished (1998)		
COMMENT	Other:GSS: CIT-HSP-2165119.TF		
Contact: Mark Adams			

Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: mtdams@tigr.org

Clones are available from Research Genetics (info@resgen.com). BAC  
end search page:  
[http://www.tigr.org/tdb/humgen/bac\\_end\\_search/bac\\_end\\_search.html](http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html)  
Seq primer: M13 Reverse  
Class: BAC ends.

## FEATURES

## SOURCE

Location/Qualifiers

1..489  
/organism="Homo sapiens"  
/db\_xref="GDB:7101197"  
/db\_xref="taxon:9606"  
/clone="2165119"  
/clone\_11b="CIT-HSP"  
/sex="Male"  
/cell\_type="Sperm"  
/note="Vector: pBelobAC11; Site\_1: HindIII; Site\_2:  
HindIII"

BASE COUNT 160 a 112 c 100 g 117 t  
ORIGIN

## Query Match

37.1%; Score 35.2; DB 17; Length 489;

Best Local Similarity 68.1%; Pred. No. 0.46; Mismatches 23; Indels 0; Gaps 0;

Matches 49; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

Oy 19 TCCTTCAGGAGAACAAAGACAGCCATTACCTGGAGAGAGACTGGCAACTGATTTTACC 78  
111 111111 1111111111 1111 11 111111 1111 111 1  
Db 167 TCCTTCAGGAGATCCAAAGACAGAGGCTACCTGGGAAAGACTAGCAATGATTTTCAC 226  
111 1111111111 1111111111 111111 111111 111111 111111  
Oy 79 CACAAGCCCAAA 90  
111 1111111111  
Db 227 CCTATGCCAAA 238  
111 1111111111

Search completed: May 2, 2003, 12:25:23  
Job time : 1098.14 secs



GenCore version 5.1.5  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 2, 2003, 11:30:30 ; Search time 38 Seconds  
(without alignments)  
766.692 Million cell updates/sec

Title: US-09-719-554-3\_COPY\_6956\_7050

Perfect score: 95  
Sequence: 1 ccgccttaccgcaagctc.....accacaagcccaactca 95

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_NM:\*  
1: /cgn2\_6/ptodata/2/ina/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/2/ina/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/2/ina/6A.COMB.seq:\*  
4: /cgn2\_6/ptodata/2/ina/6B.COMB.seq:\*  
5: /cgn2\_6/ptodata/2/ina/PCTUS.COMB.seq:\*  
6: /cgn2\_6/ptodata/2/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	94	98.9	2946	4 US-09-175-928-3	Sequence 3, Appli
2	90.4	95.2	374	1 US-08-686-878A-47	Sequence 47, Appli
3	90.4	95.2	374	1 US-08-721-489-1	Sequence 1, Appli
4	27	28.4	34185	4 US-09-545-481-3	Sequence 3, Appli
5	26.2	27.6	3234	1 US-08-264-534-31	Sequence 31, Appli
6	26.2	27.6	3234	1 US-08-083-590A-10	Sequence 10, Appli
7	26.2	27.6	3234	1 US-08-465-500-31	Sequence 31, Appli
8	26.2	27.6	3234	2 US-08-346-128-31	Sequence 31, Appli
9	26.2	27.6	3234	3 US-08-532-384-10	Sequence 10, Appli
10	26.2	27.6	3234	3 US-08-893-828-31	Sequence 31, Appli
11	26.2	27.6	4131	1 US-08-309-512-3	Sequence 3, Appli
12	26.2	27.6	4131	5 PCT-US92-08756A-3	Sequence 3, Appli
13	25.8	27.2	3002	4 US-09-369-364A-1	Sequence 1, Appli
14	25.4	26.7	2707	1 US-08-992-035A-5	Sequence 5, Appli
15	25.2	26.5	52297	4 US-09-426-436-1	Sequence 1, Appli
16	25.2	26.5	52297	4 US-08-705-557-1	Sequence 585, App
17	25	26.3	588	4 US-09-221-017B-585	Sequence 585, App
18	25	26.3	3789	1 US-07-872-644-42	Sequence 42, Appli
19	25	26.3	3789	1 US-08-297-494-42	Sequence 42, Appli
20	25	26.3	3789	1 US-08-297-510-42	Sequence 42, Appli
21	25	26.3	3789	1 US-08-479-532-42	Sequence 42, Appli
22	25	26.3	3789	1 US-08-455-526-42	Sequence 42, Appli
23	25	26.3	3789	1 US-08-455-525-42	Sequence 42, Appli
24	25	26.3	3789	3 US-09-139-491-42	Sequence 42, Appli
25	25	26.3	3789	5 PCT-US92-03222-42	Sequence 42, Appli
26	25	26.3	4131	1 US-07-872-644-38	Sequence 38, Appli
27	25	26.3	4131	1 US-08-297-494-38	Sequence 38, Appli

28	25	26.3	4131	1 US-08-297-510-38	Sequence 38, Appli
29	25	26.3	4131	1 US-08-479-532-38	Sequence 38, Appli
30	25	26.3	4131	1 US-08-455-526-38	Sequence 38, Appli
31	25	26.3	4131	1 US-08-455-525-38	Sequence 38, Appli
32	25	26.3	4131	3 US-09-139-491-38	Sequence 38, Appli
33	25	26.3	4131	5 PCT-US92-03222-38	Sequence 38, Appli
34	24.8	26.1	1574	4 US-09-173-581-12	Sequence 12, Appli
35	24.8	26.1	1574	4 US-09-420-915-12	Sequence 12, Appli
36	24.8	26.1	4661	4 US-09-221-017B-970	Sequence 970, App
37	24.6	25.9	314	4 US-09-328-111-35	Sequence 35, Appli
38	24.6	25.9	733	4 US-08-998-416-989	Sequence 989, App
39	24.6	25.9	4558	1 US-08-309-512-2	Sequence 2, Appli
40	24.6	25.9	4558	5 PCT-US92-08756A-2	Sequence 2, Appli
41	24.2	25.5	1058	4 US-08-818-112-45	Sequence 45, Appli
42	24.2	25.5	1058	4 US-08-818-111-45	Sequence 45, Appli
43	24.2	25.5	1058	4 US-09-036-536-45	Sequence 45, Appli
44	24.2	25.5	1058	4 US-09-072-596-45	Sequence 45, Appli
45	24.2	25.5	1620	1 US-07-616-022C-9	Sequence 9, Appli

## ALIGNMENTS

```
RESULT 1
US-09-175-928-3
: Sequence 3, Application US/09175928A
: Patent No. 6312921
: GENERAL INFORMATION:
: APPLICANT: Jacobs, Kenneth
: APPLICANT: McCoy, John M.
: APPLICANT: Lavalley, Edward R.
: APPLICANT: Collins-Racie, Lisa A.
: APPLICANT: Evans, Cheryl
: APPLICANT: Merberg, David
: APPLICANT: Treacy, Maurice
: APPLICANT: Mi, Sha
: APPLICANT: Genetics Institute, Inc.
: TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM
: FILE REFERENCE: 6006B, A1172A
: CURRENT APPLICATION NUMBER: US/09/175, 928A
: CURRENT FILING DATE: 1998-10-20
: NUMBER OF SEQ ID NOS: 62
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 3
: LENGTH: 2946
: TYPE: DNA
: ORGANISM: Homo sapiens
: US-09-175-928-3
Query Match          98.9%  Score 94:  DB 4:  Length 2946;
Best Local Similarity 100.0%; Pred. No. 4.2e-25;
Matches 94; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 2 CTGCTTATGCGCAAGCTCTTCAGAGACAAAGACAGGCATTACCTGGAGAGAC 61
DB 6 CTGCTTATGCGCAAGCTCTTCAGAGACAAAGACAGGCATTACCTGGAGAGAC 65
OY 62 TGGCACTGATTTTACCCCAAGCCCAACTCA 95
DB 66 TGGCACTGATTTTACCCCAAGCCCAACTCA 99
RESULT 2
US-08-686-878A-47
: Sequence 47, Application US/08686878A
: Patent No. 5708157
: GENERAL INFORMATION:
: APPLICANT: Jacobs, Kenneth
: APPLICANT: McCoy, John
: APPLICANT: Lavalley, Edward
: APPLICANT: Racie, Lisa
: APPLICANT: Merberg, David
: APPLICANT: Treacy, Maurice
```

APPLICANT: Evans, Cheryl  
APPLICANT: Spaulding, Ylki  
TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES  
TITLE OF INVENTION: ENCODING THEM  
NUMBER OF SEQUENCES: 71  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genetics Institute, Inc.  
STREET: 87 CambridgePark Drive  
CITY: Cambridge  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02140  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/686,878A  
FILING DATE:  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: Brown, Scott A.  
REGISTRATION NUMBER: 32,724  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 498-8224  
TELEFAX: (617) 876-5851  
INFORMATION FOR SEQ ID NO: 47:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 374 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
US-08-686-878A-47

Query Match 95.2%; Score 90.4; DB 1; Length 374;  
Best Local Similarity 96.8%; Pred. No. 3.9e-24;  
Matches 91; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 CTCGCTTATCGCAAGCTCTTCAGAGAAACAAGAGCCATTACCTGGAGAAGC 61  
DB 29 CTCGCTTATCGCAAGCTCTTCAGAGAAACAAGAGCCATTACCTGGAAAANAC 88

QY 62 TGGCAACTGATTTTACCACCAAGCCCAACCTCA 95  
DB 89 TGGCAACTGATTTTACCACCAAGCCCAACCTCA 122

RESULT 3  
US-08-721-489-1  
Sequence 1, Application US/08721489  
Patent No. 5786465  
GENERAL INFORMATION:  
APPLICANT: Jacobs, Kenneth  
APPLICANT: McCoy, John  
APPLICANT: Lavallee, Edward  
APPLICANT: Racie, Lisa  
APPLICANT: Merberg, David  
APPLICANT: Treacy, Maurice  
TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES  
TITLE OF INVENTION: ENCODING THEM  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genetics Institute, Inc.  
STREET: 87 CambridgePark Drive  
CITY: Cambridge  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02140  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/721,489  
FILING DATE:  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Brown, Scott A.  
REGISTRATION NUMBER: 32,724  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 498-8224  
TELEFAX: (617) 876-5851  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 374 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
US-08-721-489-1

Query Match 95.2%; Score 90.4; DB 1; Length 374;  
Best Local Similarity 96.8%; Pred. No. 3.9e-24;  
Matches 91; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 CTCGCTTATCGCAAGCTCTTCAGAGAAACAAGAGCCATTACCTGGAGAAGC 61  
DB 29 CTCGCTTATCGCAAGCTCTTCAGAGAAACAAGAGCCATTACCTGGAAAANAC 88

QY 62 TGGCAACTGATTTTACCACCAAGCCCAACCTCA 95  
DB 89 TGGCAACTGATTTTACCACCAAGCCCAACCTCA 122

RESULT 4  
US-09-545-481-3  
Sequence 3, Application US/09545481  
Patent No. 6451319  
GENERAL INFORMATION:  
APPLICANT: Chiang, Christina H.  
APPLICANT: Cochran, Mark D.  
TITLE OF INVENTION: No. 6451319el Recombinant And Mutant Adenoviruses  
FILE REFERENCE: SY0993K US  
CURRENT APPLICATION NUMBER: US/09/545,481  
CURRENT FILING DATE: 2000-04-07  
PRIOR APPLICATION NUMBER: 60/128,766  
PRIOR FILING DATE: 1999-04-09  
SOFTWARE: Patentin Ver. 2.1  
SEQ ID NO 3  
LENGTH: 34185  
TYPE: DNA  
ORGANISM: Bovine adenovirus type 1  
US-09-545-481-3

Query Match 28.4%; Score 27; DB 4; Length 34185;  
Best Local Similarity 56.0%; Pred. No. 4.6;  
Matches 51; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 1 CTCGCTTATCGCAAGCTCTTCAGAGAAACAAGAGCCATTACCTGGAGAAGA 60  
DB 24080 CCGCCATTCTCGGCCCATATACATATAAAAAAGCGGGAATAATCTTGCTTAAA 24139

QY 61 CTGGCAACTGATTTTACCACCAAGCCCAAC 91  
DB 24140 AGGGCCACGCGGTACTAGACCCCAAAAC 24170

RESULT 5  
US-08-264-534-31  
Sequence 31, Application US/08264534  
Patent No. 5648464



```

: SEQUENCE CHARACTERISTICS:
: LENGTH: 3234 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: unknown
: MOLECULE TYPE: cDNA
:
US-08-465-500-31

```

Query Match	27.6%;	Score 26.2;	DB 1;	Length 3234;
Best Local Similarity	67.3%;	Pred. NO. 3.5;		
Matches 37; Conservative	0;	Mismatches 18;	Indels 0;	Gaps 0;

0y 11 CGCCAGCTCCTTCAGAGACAAAGAACAAGGCATTACCCTGGAGAAAGCTGGC 65  
||||| ||| | |||| | | |||| | |||| | ||||  
Db 1470 CGCCAAGCGCTGTGAGGCCAGCGCATGATGCCAACATCCAGGACCAACTGGGC 1524

RESULT 8  
US-08-346-128-31

; Sequence 31, Application US/08346128  
; Patent No. 5856441

```

: GENERAL INFORMATION:
:
: APPLICANT: Artavanis-Tsakonas, Spyridon et al.
:
: TITLE OF INVENTION: Human No. 5556441c And Delta, Binding Domains
:
: TITLE OF INVENTION: In Toporythmic Proteins, And Methods Based Thereon
:
: NUMBER OF SEQUENCES: 37
:
: CORRESPONDENCE ADDRESS:

```

```

? REGISTRATION NUMBER: 18, 872
? REFERENCE/DOCKET NUMBER: 7326-009
? TELECOMMUNICATION INFORMATION:
? TELEPHONE: 212 790-9090
? TELEFAX: 212 8698864/9741
? TELETYPE: 66141 PENNIE
? INFORMATION FOR SEQ ID NO: 31:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 3234 base pairs
? TYPE: nucleic acid
? STRANDEDNESS: double
? TOPOLOGY: unknown
? MOLECULE TYPE: cDNA
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 1..3234
? OS-08-346-128-31

```

Query Match	27.6%	Score 26.2	DB 2	Length 3234
Best Local Similarity	67.3%	Pred. No. 3.5		
Matches 37; Conservative	0; Mismatches 18; Indels 0; Gaps 0			
11	CGCCAACTCCTTCAGGAGAACAAAGAACAGGCCATTACCTGGAGAGAACTGGC	65		
Db	1470 CGCCAGGCGCTGCTGGAGGCGCCAGGCGAGATGCCAACATTCAGGAGCAACATG66G	1524		

RESULT: 9  
US-08-532-384-10  
; Sequence 10, Application US/08532384  
; Patent No. 6083904  
; Inventor: INFORMATION

```

:  APPLICANT: Artavanis-Tsakonas, S. et al.
:  TITLE OF INVENTION: Therapeutic And Diagnostic Methods
:  TITLE OF INVENTION: And Compositions Based On No. 6083504ch Proteins And
:  TITLE OF INVENTION: Nucleic Acids
:  NUMBER OF SEQUENCES: 21

```

Query Match	27.6%	Score 26.2	DB 3	Length 3234
Best Local Similarity	67.38%	Pred. No. 3.5		
Matches 37; Conservative	0	Mismatches	18	Indels 0; Gaps 0

Qy 11 CGCGAAGCTCTTTCAGGAGAAACAAAGAGCCATTACCTTGGAGAGAGACTGGC 65  
|||||  
Db 1470 CGCGAAGCGCTGCTGGAGAGCCAGCGCAGATGCCAACATCCAGAGACAACATGGGC 1522

RESULT 10  
US-08-893-828-31

; Sequence 31, Application US/08893828  
; Patent No. 6090922

; APPLICANT: Artavants-Tsakonas, Spyridon  
 ; APPLICANT: Muskavitch, Marc A.T.  
 ; APPLICANT: Fehon, Richard G.  
 ; APPLICANT: Rebay, Ilaria  
 ; APPLICANT: Blummueller, Cristtine M.

TITLE OF INVENTION:	HUMAN NOTCH AND DELTA, BINDING DOMAINS
TITLE OF INVENTION:	IN TOPOGRAPHIC PROTEINS, AND METHODS BASED THEREON
NUMBER OF SEQUENCES:	34



TELEFAX: 415-433-8716  
TELEX: 278356  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 4131 base pairs  
TYPE: NUCLEIC ACID  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Acetobacter xylinum  
PCT-US92-08735A-3

Query Match 27.6%; Score 26.2; DB 5; Length 4131;  
Best Local Similarity 60.6%; Pred. No. 3.9;  
Matches 43; Conservative 0; Mismatches 28; Indels 0; Gaps 0;

QY 2 CTGCGTTATCGCCAGGCTCCTTCAGAGAACAAAGAGCCATTACCTGAGAGAGAC 61  
DB 2317 CAGCTTCATCAACGATTCGAGCATGACACCAATGCCAGGCCGTGACCATGGCGTGAT 2376

QY 62 TGGCACTGAT 72  
DB 2377 CGGCATCGCT 2387

RESULT 13  
US-09-369-364A-1  
Sequence 1, Application US/09369364A  
Patent No. 6391610  
GENERAL INFORMATION:  
APPLICANT: Apte, Suneel  
APPLICANT: Hurskainen, Tiina L.  
APPLICANT: Hirohata, Satoshi  
TITLE OF INVENTION: Nucleic Acids Encoding Zinc Metalloproteases  
FILE REFERENCE: 26473/4007/10-30-00  
CURRENT APPLICATION NUMBER: US/09/369,364A  
CURRENT FILING DATE: 1999-08-06  
NUMBER OF SEQ ID NOS: 31  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 1  
LENGTH: 3002  
TYPE: DNA  
ORGANISM: mus musculus ADAMTS-5  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (18)..(2810)  
US-09-369-364A-1

Query Match 27.2%; Score 25.8; DB 4; Length 3002;  
Best Local Similarity 58.4%; Pred. No. 4.8;  
Matches 45; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 12 GCCAGCTCCTTCAGGAGAACAAAGAGCCATTACCTGAGAGAGACTGGCACTGA 71  
DB 990 GACACGAGTGTGAGGAGACAAATGCGCCGACGACCTCAAGAACTTTGCAATGG 1049

QY 72 TTTTACCACAGCCCA 88  
DB 1050 CAGCACCACATACCA 1066

RESULT 14  
US-08-992-035A-5/C  
Sequence 5, Application US/08992035A  
Patent No. 6242179  
GENERAL INFORMATION:  
APPLICANT: Shah, Puryi  
APPLICANT: Hillman, Jennifer L.  
APPLICANT: Corley, Neil C.  
APPLICANT: Lal, Preeti

TITLE OF INVENTION: HUMAN PHOSPHATASES  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Incyte Pharmaceuticals, Inc.  
STREET: 3174 Porter Dr.  
CITY: Palo Alto  
STATE: CA  
COUNTRY: USA  
ZIP: 94304

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/992,035A  
FILING DATE: December 17, 1997  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Billings, Lucy J.

REGISTRATION NUMBER: 36,749  
REFERENCE/DOCKET NUMBER: PF-0433 US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 650-845-0555  
TELEFAX: 650-845-4166

INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2707 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear

IMMEDIATE SOURCE:  
LIBRARY: COLN00722  
CLONE: 1734452  
US-08-992-035A-5

Query Match 26.7%; Score 25.4; DB 4; Length 2707;  
Best Local Similarity 61.2%; Pred. No. 6.5;  
Matches 41; Conservative 0; Mismatches 26; Indels 0; Gaps 0;

QY 21 CTCGAGAGACAAAGACAGCCATTACCTGAGAGAGCTGGCAACTGTTTACCA 80  
DB 2553 CTTGGGATCAAAAGACCAAGCCAGCTCCCTGCGATGGCTGTGTAACCA 2494

QY 81 CAAGCC 87  
DB 2493 CAAGCC 2487

RESULT 15  
US-09-426-436-1/C  
Sequence 1, Application US/09426436  
Patent No. 6225066  
GENERAL INFORMATION:  
APPLICANT: William R. Jacobs, Jr.  
APPLICANT: Barry R. Bloom  
APPLICANT: Graham F. Hatfull  
TITLE OF INVENTION: MYCOBACTERIAL SPECIES-SPECIFIC  
NUMBER OF SEQUENCES: 1  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Amster, Rothstein & Ebenstein  
STREET: 90 Park Avenue  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10016  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch 1.44 Mb storage diskette  
COMPUTER: IBM PC Compatible  
OPERATING SYSTEM: MS-DOS

```
SOFTWARE: Word Processor (ASCII)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/426,436
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/705,557
FILING DATE:
APPLICATION NUMBER: US/08/057,531
FILING DATE:
APPLICATION NUMBER: 07/833,431
FILING DATE: February 7, 1992
ATTORNEY/AGENT INFORMATION:
NAME: Pasqualini, Patricia A.
REGISTRATION NUMBER: 34,894
REFERENCE/DOCKET NUMBER: 96700/238
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 697-5995
TELEFAX: (212) 286-0854 or 286-0082
TELEX: TWX 710-581-4766
INFORMATION FOR SEQ. ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 52297
TYPE: nucleotide
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE:
DESCRIPTION: phage genome sequence
ANTHETICAL: no
ANTI-SENSE: no
FRAGMENT TYPE: not applicable.
ORIGINAL SOURCE:
ORGANISM: mycobacteriophage L5
STRAIN: not applicable
INDIVIDUAL ISOLATE: L5
DEVELOPMENTAL STAGE: not applicable
HAPLOTYPE: not applicable
TISSUE TYPE: not applicable
CELL TYPE: not applicable
ORGANELLE: not applicable
IMMEDIATE SOURCE: mycobacteriophage L5 particles
POSITION IN GENOME: entire genome
FEATURE:
NAME/KEY:
LOCATION:
IDENTIFICATION METHOD:
OTHER INFORMATION:
PUBLICATION INFORMATION:
AUTHORS: Hatfull and Sarkis
TITLE: DNA Sequence, Structure and Gene
Expression of Mycobacteriophage L5:
TITLE: A Phage System for Mycobacterial
Genetics
JOURNAL: Molecular Microbiology
VOLUME: 7
PAGES: 385-405
DATE: 1993
US-09-426-436-1

Query Match 26.5%; Score 25.2; DB 4; Length 52297;
Best Local Similarity 66.7%; Pred. No. 25;
Matches 36; Conservative 0; Mismatches 18; Indels 0; Gaps 0;
```

```
OY 10 TCGCAGCTCTCTCAGGAGCAAGCAAGCATTACCTGAGAGACTG 63
|||||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 47011 TCGCAGCTCTCTCGCGCAGCGCAGAGCTGCGCGTACCGCGAGAGGCTG 46958
```

Search completed: May 2, 2003, 12:49:30  
Job time : 56 secs





GenCore version 5.1.5  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 2, 2003, 12:25:30 ; Search time 63.2241 Seconds

(without alignments)  
1775.493 Million cell updates/sec

Title: US-09-719-554-3\_COPY\_6956\_7050

Perfect score: 95  
Sequence: 1 ccgtccctatgcacgaagc.....accacacagcccaactca 95

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 746064 seqs, 590810554 residues

Total number of hits satisfying chosen parameters: 1492128

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published\_Applications\_NA:\*  
1: /cgn2\_6/prodata/1/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/prodata/1/pubpna/PCR\_NEW\_PUB.seq:\*  
3: /cgn2\_6/prodata/1/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/prodata/1/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/prodata/1/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/prodata/1/pubpna/PCRUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/prodata/1/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/prodata/1/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/prodata/1/pubpna/US09\_NEW\_PUB.seq:\*  
10: /cgn2\_6/prodata/1/pubpna/US09\_PUBCOMB.seq:\*  
11: /cgn2\_6/prodata/1/pubpna/US10\_NEW\_PUB.seq:\*  
12: /cgn2\_6/prodata/1/pubpna/US10\_PUBCOMB.seq:\*  
13: /cgn2\_6/prodata/1/pubpna/US60\_NEW\_PUB.seq:\*  
14: /cgn2\_6/prodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	94	98.9	2930	10	US-09-902-535-1
2	94	98.9	2946	9	US-10-114-893-134
3	90.4	95.2	374	12	US-10-040-916-47
4	81.2	85.5	337	10	US-09-867-701-3363
5	71.2	74.9	758	7	US-08-979-847-112
6	30.8	32.4	414	10	US-09-864-761-25108
7	30.8	32.4	565	10	US-09-864-761-8377
8	29.4	30.9	222	10	US-09-974-300-8046
9	28.6	30.1	78056	9	US-10-109-551-1
10	28.6	30.1	1503841	9	US-09-946-807-1
11	28.6	30.1	1503841	10	US-09-795-668-1
12	28.6	30.1	1503841	10	US-09-795-668-1
13	27.8	29.3	3657	10	US-09-728-422-4
14	27.6	29.1	15070	10	US-09-764-877-3979
15	27.6	29.1	15071	10	US-09-764-877-3978
16	26.8	28.2	519	9	US-10-091-572-130
17	26.8	28.2	519	9	US-09-764-891-2099
18	26.8	28.2	8833	10	US-09-880-107-3272
19	26.2	27.6	7332	10	US-09-944-849-1

20	26.2	27.6	7673	10	US-09-815-925-1	Sequence 1, Appl1
21	26	27.4	1965	10	US-09-940-921B-3	Sequence 3, Appl1
22	26	27.4	2027	10	US-09-867-550-887	Sequence 887, App
23	26	27.4	2052	10	US-09-940-921B-1	Sequence 1, Appl1
24	26	27.4	2240	10	US-09-940-921B-5	Sequence 5, Appl1
25	26	27.4	2311	10	US-09-925-297-94	Sequence 94, Appl1
26	25.8	27.2	400	10	US-09-878-574-3631	Sequence 3631, Ap
27	25.8	27.2	3002	10	US-09-918-171A-1	Sequence 1, Appl1
28	25.8	27.2	5938	10	US-09-954-456-1607	Sequence 1607, Ap
29	25.8	27.2	5938	10	US-09-967-768A-225	Sequence 225, App
30	25.6	26.9	302	10	US-09-878-574-14590	Sequence 14590, A
31	25.4	26.7	329	10	US-09-960-352-8386	Sequence 8386, Ap
32	25.4	26.7	417	10	US-09-960-352-13655	Sequence 13655, A
33	25.4	26.7	450	9	US-09-918-995-25995	Sequence 25995, A
34	25.4	26.7	696	10	US-09-728-446-115	Sequence 115, App
35	25.4	26.7	859	10	US-09-925-300-683	Sequence 683, App
36	25.4	26.7	1939	9	US-09-925-299-89	Sequence 89, Appl
37	25.4	26.7	1939	10	US-09-925-299-89	Sequence 89, Appl
38	25.4	26.7	2272	9	US-10-174-590-345	Sequence 345, App
39	25.4	26.7	2272	9	US-10-176-758-345	Sequence 345, App
40	25.4	26.7	2272	9	US-10-175-737-345	Sequence 345, App
41	25.4	26.7	2272	9	US-10-173-706-345	Sequence 345, App
42	25.4	26.7	2272	9	US-10-175-738-345	Sequence 345, App
43	25.4	26.7	2272	9	US-10-175-752-345	Sequence 345, App
44	25.4	26.7	2272	9	US-10-176-482-345	Sequence 345, App
45	25.4	26.7	2272	9	US-10-176-757-345	Sequence 345, App

## ALIGNMENTS

RESULT 1  
US-09-902-535-1  
Sequence 1, Application US/09902535  
Patent No. US20020102530A1  
GENERAL INFORMATION:  
APPLICANT: Keith, Jr., James C.  
APPLICANT: McCoy, John M.  
TITLE OF INVENTION: Methods and compositions for diagnosing  
TITLE OF INVENTION: and treating preclampsia and gestational trophoblast  
FILE REFERENCE: GIN-6006B4  
CURRENT APPLICATION NUMBER: US/09/902,535  
CURRENT FILING DATE: 2001-07-09  
PRIOR APPLICATION NUMBER: 60/216,657  
PRIOR FILING DATE: 2000-07-06  
NUMBER OF SEQ ID NOS: 5  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 1  
LENGTH: 2930  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (930)...(2546)  
US-09-902-535-1  
Query Match 98.9%; Score 94; DB 10; Length 2930;  
Best Local Similarity 100.0%; Pred. No. 3.5e-25;  
Matches 94; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 2 CTGCTTATGCCCAAGCTCTTCAGAGAAACAAGAGCCATTACCTGGAGAAC 61  
|||||  
DB 8 CTGCTTATGCCCAAGCTCTTCAGAGAAACAAGAGCCATTACCTGGAGAAC 67  
|||||  
QY 62 TGGCACTGATTATCCCAAGCCCAAGCCCAAGCTCA 95  
|||||  
DB 68 TGGCACTGATTATCCCAAGCCCAAGCCCAAGCTCA 101  
|||||  
RESULT 2  
US-10-114-893-134

Sequence 134, Application US/10114893  
Publication No. US20020193567A1  
GENERAL INFORMATION:  
APPLICANT: Jacobs, Kenneth  
APPLICANT: McCoy, John M.  
APPLICANT: Lavallee, Edward R.  
APPLICANT: Collins-Racie, Lisa A.  
APPLICANT: Evans, Cheryl  
APPLICANT: Merberg, David  
APPLICANT: Treacy, Maurice  
APPLICANT: Bowman, Michael R.  
APPLICANT: Spaulding, Vikki  
APPLICANT: Carlin-Duckett, McKenough  
APPLICANT: Kelleher, Kerry S.  
APPLICANT: Genetics Institute, Inc.  
TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM  
FILE REFERENCE: GI 6000-10A  
CURRENT APPLICATION NUMBER: US/10/114,893  
CURRENT FILING DATE: 2002-04-02  
EARLIER APPLICATION NUMBER: 09/413,232  
EARLIER FILING DATE: 1999-10-06  
NUMBER OF SEQ ID NOS: 321  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 134  
LENGTH: 2946  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-10-114-893-134

Query Match 98.9%; Score 94; DB 9; Length 2946;  
Best Local Similarity 100.0%; Pred. No. 3.5e-25;  
Matches 94; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 CTGCTTATCGCAAGCTCTTCAGAGAAACAAGAGCCATTACCTGGAGAAGAC 61  
DB 6 CTGCTTATCGCAAGCTCTTCAGAGAAACAAGAGCCATTACCTGGAGAAGAC 65  
QY 62 TGGCAACTGATTTTACCCACAAAGCCCAAACTCA 95  
DB 66 TGGCAACTGATTTTACCCACAAAGCCCAAACTCA 99

RESULT 3  
US-10-040-916-47  
Sequence 47, Application US/10040916  
Patent No. US20020146769A1  
GENERAL INFORMATION:  
APPLICANT: Jacobs, Kenneth  
McCoy, John  
Lavallee, Edward  
Racie, Lisa  
Merberg, David  
Treacy, Maurice  
Evans, Cheryl  
Spaulding, Vikki

TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES  
ENCODING THEM

NUMBER OF SEQUENCES: 71  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genetics Institute, Inc.  
STREET: 87 Cambridgepark Drive  
CITY: Cambridge  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02140

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA: US/10/040,916  
FILING DATE: 07-Jan-2002

CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/887,029  
FILING DATE: 07-FEB-1997  
APPLICATION NUMBER: 08/686,878  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Brown, Scott A.  
REGISTRATION NUMBER: 32,724  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 498-8224  
TELEFAX: (617) 876-5851  
INFORMATION FOR SEQ ID NO: 47:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 374 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
SEQUENCE DESCRIPTION: SEQ ID NO: 47:  
US-10-040-916-47

Query Match 95.2%; Score 90.4; DB 12; Length 374;  
Best Local Similarity 96.8%; Pred. No. 4.1e-24;  
Matches 91; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 2 CTGCTTATCGCAAGCTCTTCAGAGAAACAAGAGCCATTACCTGGAGAAGAC 61  
DB 29 CTGCTTATCGCAAGCTCTTCAGAGAAACAAGAGCCATTACCTGGAAAAAANC 88  
QY 62 TGGCAACTGATTTTACCCACAAAGCCCAAACTCA 95  
DB 89 TGGCAACTGATTTTACCCACAAAGCCCAAACTCA 122

RESULT 4  
US-09-867-701-3363/C  
Sequence 3363, Application US/09867701  
Patent No. US2002013237A1  
GENERAL INFORMATION:  
APPLICANT: Aglate, Paul A.  
APPLICANT: Jones, Robert  
APPLICANT: Harlocker, Susan L.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY  
TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER  
FILE REFERENCE: 210121.497  
CURRENT APPLICATION NUMBER: US/09/867,701  
CURRENT FILING DATE: 2001-05-29  
NUMBER OF SEQ ID NOS: 10912  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 3363  
LENGTH: 337  
TYPE: DNA  
ORGANISM: Homo sapien  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)...(337)  
OTHER INFORMATION: n = A,T,C or G  
US-09-867-701-3363

Query Match 85.5%; Score 81.2; DB 10; Length 337;  
Best Local Similarity 90.5%; Pred. No. 1.1e-20;  
Matches 86; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 1 CCGCTTATCGCAAGCTCTTCAGAGAAACAAGAGCCATTACCTGGAGAAGAC 60  
DB 184 CCGCTTATCGCAAGCTCTTCAGAGAAACAAGAGCCATTACCTGGAGAAGAC 125  
QY 61 CTGGCAACTGATTTTACCCACAAAGCCCAAACTCA 95  
DB 124 CTGGCAACTGATTTTACCCACAAAGCCCAAACTCA 90

```
RESULT 5
US-08-979-847-112
; Sequence 112, Application US/08979847
; Publication No. US20030039664A1
; GENERAL INFORMATION:
; APPLICANT: PERRON, HERVE
; APPLICANT: BESEME, FREDERIC
; APPLICANT: BEDI, FREDERIC
; APPLICANT: PARANHOS-BACCALA, GLAUCIA
; APPLICANT: KOMORIAN-PRADEL, FLORENCE
; APPLICANT: JOLIVET-REYNAUD, COLETTE
; APPLICANT: MANDRAND, BERNARD
; APPLICANT: GARSON, JEREMY
; APPLICANT: TURE, PHILIP
; TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS
; TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACT
; TITLE OF INVENTION: THERAPEUTIC PURPOSES
; NUMBER OF SEQUENCES: 210
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: OLIVE & BERRIDGE, PLC
; STREET: P.O. BOX 19928
; CITY: ALEXANDRIA
; STATE: VA
; COUNTRY: USA
; ZIP: 22320
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/979,847
; FILING DATE: 26-NOV-1997
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: BERRIDGE, WILLIAM P.
; REGISTRATION NUMBER: 30,024
; REFERENCE/DOCKET NUMBER: WPB 39046A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703-836-6400
; TELEFAX: 703-836-2787
; INFORMATION FOR SEQ ID NO: 112:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 758 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-979-847-112

Query Match          74.9%; Score 71.2; DB 7; Length 758;
Best Local Similarity 90.6%; Pred. No. 7.6e-17;
Matches 8; Conservative 0; Mismatches 8; Indels 1; Gaps 1;

QY 1 CCGGCTTATGCGCAAGCTCTTCAGAGAAAGAAAGAGCCATTACCTGGAGAGA 60
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 107 CCGGCTTATGCGCATCTTCCTTCAGAGAAAGAAAGAGCCATTACCGAGGAAGA 166
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 61 CTGGCAACT-GATTTTACCAACAAGCCCAACTCA 95
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 167 CTGGCAACTGATTTTACCAACATGCGCAATGTC 202
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 6
US-09-864-761-25108
; Sequence 25108, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
```

```
; TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY
; FILE REFERENCE: Aemica-x-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263.6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 25108
; LENGTH: 414
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AC019191.2
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 4.4
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.6
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.7
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.2
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.4
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.9
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.4
; OTHER INFORMATION: SWISSPROT HIT: P21414, EVALU6 3.00e-16
; OTHER INFORMATION: NT HIT: Z78766.1, EVALU6 4.00e-62
; OTHER INFORMATION: EST_HUMAN HIT: BE142076.1, EVALU6 1.00e-59
US-09-864-761-25108

Query Match          32.4%; Score 30.8; DB 10; Length 414;
Best Local Similarity 61.0%; Pred. No. 0.077;
Matches 50; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 9 ATCGCAAGCTCTTCAGAGAAAGAAAGAGCCATTACCTGGAGAGACTGGCAAC 68
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 152 ATCATAGCGCATCTCGGGGAAACAAAGAAATATGCACTATCCCTGAGAGAACTGGCACT 211
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 69 TGATTTACCAACAAGCCCAAA 90
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 212 TAGACTTCACCAATATGCTTAA 233
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```

RESULT 7
US-09-864-761-8377
: Sequence 8377, Application US/09864761
: Patent No. US20020048763A1
: GENERAL INFORMATION:
: APPLICANT: Penn, Sharon G.
: APPLICANT: Rank, David R.
: APPLICANT: Hanzel, David K.
: APPLICANT: Chen, Wensheng
: TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
: FILE REFERENCE: Aecomica-X-1
: CURRENT FILING DATE: US/09/864,761
: PRIOR APPLICATION NUMBER: US 60/180,312
: PRIOR FILING DATE: 2000-02-04
: PRIOR APPLICATION NUMBER: US 60/207,456
: PRIOR FILING DATE: 2000-05-26
: PRIOR APPLICATION NUMBER: US 09/632,366
: PRIOR FILING DATE: 2000-08-03
: PRIOR APPLICATION NUMBER: GB 24263.6
: PRIOR FILING DATE: 2000-10-04
: PRIOR APPLICATION NUMBER: US 60/236,359
: PRIOR FILING DATE: 2000-09-27
: PRIOR APPLICATION NUMBER: PCT/US01/00666
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00667
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00664
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00669
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00665
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00668
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00663
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00662
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00661
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: PCT/US01/00670
: PRIOR FILING DATE: 2001-01-30
: PRIOR APPLICATION NUMBER: US 60/234,687
: PRIOR FILING DATE: 2000-09-21
: PRIOR APPLICATION NUMBER: US 09/608,408
: PRIOR FILING DATE: 2000-06-30
: PRIOR APPLICATION NUMBER: US 09/774,203
: PRIOR FILING DATE: 2001-01-29
: NUMBER OF SEQ ID NOS: 49117
: SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
: SEQ ID NO 8377
: LENGTH: 565
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: OTHER INFORMATION: MAP TO AC019191.2
: OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 4.4
: OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.6
: OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 2.7
: OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3
: OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.2
: OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2.4
: OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.9
: OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3.4
: US-09-864-761-8377
Query Match 32.4%; Score 30.8; DB 10; Length 565;
Best Local Similarity 61.0%; Pred. No. 0.085;
Matches 50; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

```

```
QY      9  ATCGGCAAGCTCCTTCAGAGAACAAGACGGCCATTACCCTGGAGAACTGGCAAC   68
        ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     250  ATCATTAAGGCCACTCTGGGGAACAAGAATATGGCACTATCTGGAGCAACTGGCACT   309
        | . | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY      69  TGATTTTACCACAGCCCAA   90
        | . | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db     310  TAGACTTGACCCATTATGCCTAA   331

RESULT 8
US-09-974-300-8046
; Sequence 8046, Application US/09974300
; Patent No. US2002014672A1
GENERAL INFORMATION:
; APPLICANT: Berka, Randy M.
; TITLE OF INVENTION: Methods For Monitoring Multiple Gene
; FILE REFERENCE: 10085-500-US
CURRENT APPLICATION NUMBER: US/09/974,300
CURRENT FILING DATE: 2001-10-05
PRIOR APPLICATION NUMBER: 09/680,598
PRIOR FILING DATE: 2000-10-06
PRIOR APPLICATION NUMBER: 60/279,526
PRIOR FILING DATE: 2001-03-27
NUMBER OF SEQ ID NOS: 8481
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 8046
LENGTH: 222
TYPE: DNA
ORGANISM: Bacillus clausii
US-09-974-300-8046

Query Match          30.9%; Score 29.4; DB 10; Length 222;
Best Local Similarity 60.8%; Pred.No.0.21;
Matches 48; Conservative 0; Mismatches 31; Indels 0; Gaps 0.

QY      5  CCTTATGCCAAGCTCCTTCAGAGAACAAGACGGCCATTACCCTGGAGAACTGG   64
        ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db     27  CCTCATCTCCCAAGCCCTTAATGAGAACGACGACCATCTCGCTTCACGAACTAGATGA   86
        |||| |||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY      65  CAACGTATTTTACCCCAA   83
        |||| |||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db     87  CAACGATTAAAGCGCAA   105

RESULT 9
US-10-109-551-1
; Sequence 1, Application US/10109551
; Publication No. US20020194635A1
GENERAL INFORMATION:
; APPLICANT: DUNNE, PATRICK W.
; APPLICANT: PIEDRAHITA, JORGE
; TITLE OF INVENTION: TRANSGENIC ANIMALS RESISTANT TO TRANSMISSIBLE
; FILE REFERENCE: TARK:207US
CURRENT APPLICATION NUMBER: US/10/109,551
CURRENT FILING DATE: 2002-03-28
PRIOR APPLICATION NUMBER: 60/280,549
PRIOR FILING DATE: 2001-03-30
NUMBER OF SEQ ID NOS: 10
SOFTWARE: Patentin Ver. 2.1
SEQ ID NO 1
LENGTH: 78056
TYPE: DNA
ORGANISM: Bos taurus
US-10-109-551-1

Query Match          30.1%; Score 28.6; DB 9; Length 78056;
Best Local Similarity 61.3%; Pred.No.2.6;
Matches 46; Conservative 0; Mismatches 29; Indels 0; Gaps 0.

QY      16  AGCTCTTCAGAGACAACAAGACAGGCCATTACCCCTGSAGAGACTGGCAACTGATTTT   75
```

Db 29105 AGCAAAATCGGTACATGAGTGAATATCTGAGAGACTGGAAATTGATTT 29164

QY 76 ACCCAACAAGCCCAAA 90

Db 29165 TACTCATATGCCAAA 29179

## RESULT 10

US-09-946-807-1/c  
Sequence 1, Application US/09946807  
Patent No. US20020165144A1  
GENERAL INFORMATION:  
APPLICANT: Stefansson, Hreinn  
APPLICANT: Steinthorsdottir, Valgerdur  
APPLICANT: Gulcher, Jeffrey R.  
TITLE OF INVENTION: HUMAN SCHIZOPHRENIA GENE  
FILE REFERENCE: 2345.2004-001  
CURRENT APPLICATION NUMBER: US/09/946,807  
CURRENT FILING DATE: 2001-09-05  
PRIORITY APPLICATION NUMBER: US/09/795,668  
PRIORITY FILING DATE: 2001-02-28  
PRIORITY APPLICATION NUMBER: US 09/515,716  
PRIORITY FILING DATE: 2000-02-28  
NUMBER OF SEQ ID NOS: 1531  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 1  
LENGTH: 1503841  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)...(1531)  
OTHER INFORMATION: y=t/u or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)...(1531)  
OTHER INFORMATION: r=g or a  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)...(1531)  
OTHER INFORMATION: m=a or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)...(1531)  
OTHER INFORMATION: k=g or t/u  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)...(1531)  
OTHER INFORMATION: s=g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)...(1531)  
OTHER INFORMATION: w=a or t/u  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)...(1531)  
OTHER INFORMATION: d=a or g or t/u  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)...(1531)  
OTHER INFORMATION: h=a or c or t/u  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)...(1531)  
OTHER INFORMATION: v=a or g or c  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)...(1531)

OTHER INFORMATION: n=a or g or c or t/u  
US-09-946-807-1

Query Match 30.1%: Score 28.6; DB 9; Length 1503841;  
Best Local Similarity 64.2%: Pred. No. 6.4;  
Matches 43; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

QY 13 CCAAGCTCCCTCAGAGAAACAAGACGACCATACCTGGAGAGACTGGCAACTGAT 72

Db 784413 CGAAATCCTTCAGTAGAAGAAACAATGTAATTCTCGAGAGATGCTCCCTTCTGT 784354

QY 73 TTATCCC 79

Db 784353 ATTGCC 784347

## RESULT 11

US-09-795-668-1/c  
Sequence 1, Application US/09795668  
Patent No. US20020045577A1  
GENERAL INFORMATION:  
APPLICANT: Stefansson, Hreinn  
APPLICANT: Steinthorsdottir, Valgerdur  
APPLICANT: Gulcher, Jeffrey R.  
TITLE OF INVENTION: HUMAN SCHIZOPHRENIA GENE  
FILE REFERENCE: 2345.2004-001  
CURRENT APPLICATION NUMBER: US/09/795,668  
CURRENT FILING DATE: 2001-02-28  
PRIORITY APPLICATION NUMBER: US 09/515,716  
PRIORITY FILING DATE: 2000-02-28  
NUMBER OF SEQ ID NOS: 1531  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 1  
LENGTH: 1503841  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)...(1531)  
OTHER INFORMATION: r=g or a  
NAME/KEY: misc\_feature  
LOCATION: (1)...(1531)  
OTHER INFORMATION: y=t/u or c  
NAME/KEY: misc\_feature  
LOCATION: (1)...(1531)  
OTHER INFORMATION: m=a or c  
NAME/KEY: misc\_feature  
LOCATION: (1)...(1531)  
OTHER INFORMATION: k=g or t/u  
NAME/KEY: misc\_feature  
LOCATION: (1)...(1531)  
OTHER INFORMATION: s=g or c  
NAME/KEY: misc\_feature  
LOCATION: (1)...(1531)  
OTHER INFORMATION: w=a or t/u  
NAME/KEY: misc\_feature  
LOCATION: (1)...(1531)  
OTHER INFORMATION: d=a or g or t/u  
NAME/KEY: misc\_feature  
LOCATION: (1)...(1531)  
OTHER INFORMATION: h=a or c or t/u  
NAME/KEY: misc\_feature  
LOCATION: (1)...(1531)  
OTHER INFORMATION: v=a or g or c  
NAME/KEY: misc\_feature  
LOCATION: (1)...(1531)  
OTHER INFORMATION: n=a or g or c or t/u  
US-09-795-668-1

Query Match 30.1%: Score 28.6; DB 10; Length 1503841;

Best Local Similarity 64.2%; Pred. No. 6.4;  
Matches 43; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

QY 13 CCAAGCTCTTCAGGAGAACAAAGACAGCCATTACCTGGAGAGACTGGCACTGAT 72

DB 784413 CGAAATCTTCAGTAGTAGAAACAATGTAATTTCTGGAGAGATCTCCCTCTGT 784354

QY 73 TTATCCC 79

DB 784353 ATTTCCC 784347

RESULT 12

US-09-795-686-1/c  
Sequence 1, Application US/09795686

Patent No. US20020094954A1

GENERAL INFORMATION:

APPLICANT: Stefansson, Hreinn

APPLICANT: Steinthorsson, Valgerdur

APPLICANT: Gulcher, Jeffrey R.

TITLE OF INVENTION: HUMAN SCHIZOPHRENIA GENE

FILE REFERENCE: 2345-2005-001

CURRENT APPLICATION NUMBER: US/09/795,686

PRIOR FILING DATE: 2001-02-28

PRIOR APPLICATION NUMBER: US 09/515,715

NUMBER OF SEQ ID NOS: 1531

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 1

LENGTH: 1503841

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: misc\_feature

LOCATION: (1)...(1531)

NAME/KEY: misc\_feature

LOCATION: (1)...(1531)

NAME/KEY: misc\_feature

LOCATION: (1)...(1531)

NAME/KEY: misc\_feature

LOCATION: (1)...(1531)

NAME/KEY: misc\_feature

LOCATION: (1)...(1531)

NAME/KEY: misc\_feature

LOCATION: (1)...(1531)

NAME/KEY: misc\_feature

LOCATION: (1)...(1531)

NAME/KEY: misc\_feature

LOCATION: (1)...(1531)

NAME/KEY: misc\_feature

LOCATION: (1)...(1531)

NAME/KEY: misc\_feature

LOCATION: (1)...(1531)

NAME/KEY: misc\_feature

LOCATION: (1)...(1531)

NAME/KEY: misc\_feature

LOCATION: (1)...(1531)

NAME/KEY: misc\_feature

LOCATION: (1)...(1531)

NAME/KEY: misc\_feature

LOCATION: (1)...(1531)

NAME/KEY: misc\_feature

LOCATION: (1)...(1531)

NAME/KEY: misc\_feature

LOCATION: (1)...(1531)

NAME/KEY: misc\_feature

LOCATION: (1)...(1531)

NAME/KEY: misc\_feature

LOCATION: (1)...(1531)

NAME/KEY: misc\_feature

LOCATION: (1)...(1531)

NAME/KEY: misc\_feature

LOCATION: (1)...(1531)

NAME/KEY: misc\_feature

LOCATION: (1)...(1531)

NAME/KEY: misc\_feature

LOCATION: (1)...(1531)

NAME/KEY: misc\_feature

LOCATION: (1)...(1531)

Query Match 30.1%; Score 28.6; DB 10; Length 1503841;

Best Local Similarity 64.2%; Pred. No. 6.4;  
Matches 43; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

QY 13 CCAAGCTCTTCAGGAGAACAAAGACAGCCATTACCTGGAGAGACTGGCACTGAT 72

DB 784413 CGAAATCTTCAGTAGTAGAAACAATGTAATTTCTGGAGAGATCTCCCTCTGT 784354

QY 73 TTATCCC 79

DB 784353 ATTTCCC 784347

RESULT 13

US-09-728-422-4  
Sequence 4, Application US/09728422

Patent No. US20020128187A1

GENERAL INFORMATION:

APPLICANT: Tang, Y. Tom

APPLICANT: Zhou, Ping

APPLICANT: Goodrich, Ryle

APPLICANT: Liu, Chenghua

APPLICANT: Asundi, Vinod

APPLICANT: Ren, Feiyan

APPLICANT: Zhao, Qing A.

APPLICANT: Yang, Yondhong

APPLICANT: Wehman, Tom

APPLICANT: Drmanac, Radoje T.

TITLE OF INVENTION: NO. US20020128187A1el Nucleic Acids and

FILE REFERENCE: 787CIP2F

CURRENT APPLICATION NUMBER: US/09/728,422

PRIOR FILING DATE: 2000-11-30

PRIOR APPLICATION NUMBER: 09/560,875

PRIOR FILING DATE: 2000-04-27

PRIOR APPLICATION NUMBER: 09/496,914

PRIOR FILING DATE: 2000-02-03

NUMBER OF SEQ ID NOS: 10

SOFTWARE: PL-Fl-Genes Version 2.0

SEQ ID NO 4

LENGTH: 3657

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: CDS

LOCATION: (645)..(3254)

NAME/KEY: CDS

LOCATION: (645)..(3254)

NAME/KEY: CDS

LOCATION: (645)..(3254)

NAME/KEY: CDS

LOCATION: (645)..(3254)

NAME/KEY: CDS

LOCATION: (645)..(3254)

NAME/KEY: CDS

LOCATION: (645)..(3254)

NAME/KEY: CDS

LOCATION: (645)..(3254)

NAME/KEY: CDS

LOCATION: (645)..(3254)

NAME/KEY: CDS

LOCATION: (645)..(3254)

NAME/KEY: CDS

LOCATION: (645)..(3254)

NAME/KEY: CDS

LOCATION: (645)..(3254)

NAME/KEY: CDS

LOCATION: (645)..(3254)

NAME/KEY: CDS

LOCATION: (645)..(3254)

NAME/KEY: CDS

LOCATION: (645)..(3254)

NAME/KEY: CDS

LOCATION: (645)..(3254)

NAME/KEY: CDS

LOCATION: (645)..(3254)

NAME/KEY: CDS

LOCATION: (645)..(3254)

NAME/KEY: CDS

LOCATION: (645)..(3254)

NAME/KEY: CDS

LOCATION: (645)..(3254)

NAME/KEY: CDS

LOCATION: (645)..(3254)

Query Match 29.3%; Score 27.8; DB 10; Length 3657;

Best Local Similarity 59.5%; Pred. No. 2;  
Matches 47; Conservative 0; Mismatches 32; Indels 0; Gaps 0;

QY 9 ATGCCCACTCTTCAGGAGAACAAAGACAGCCATTACCTGGAGAGACTGGCACTGAT 68

DB 744 ATCTTAAGATCAATGATACAGAGATGAGATGAGATGAGATGAGATGAGATGAGAT 803

QY 69 TGATTTTACCAAGGCC 87

DB 804 TTGTTAGCAACAAGTCC 822

RESULT 14

US-09-764-877-3979/c  
Sequence 3979, Application US/09764877

Patent No. US20020147140A1

GENERAL INFORMATION:

APPLICANT: Rosen et al.

TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

FILE REFERENCE: PC005

CURRENT APPLICATION NUMBER: US/09/764,877

PRIOR FILING DATE: 2001-01-17

PRIOR APPLICATION data removed - refer to PALM or file wrapper

NUMBER OF SEQ ID NOS: 4031

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 3979

LENGTH: 15070

TYPE: DNA

ORGANISM: Homo sapiens

US-09-764-877-3979

Query Match 29.1%; Score 27.6; DB 10; Length 15070;  
Best Local Similarity 63.6%; Pred. No. 3.7;  
Matches 42; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

QY 25 AGGAGAACAAAGACAGCCATTACCTGGAGAGACTGGCACTGATTTTACCCACAAG 84

Db 6635 AGAAAAAGAAAAAGAAAAAGATATGAAACAGACAGACACATCATGTGGCCCAAAAT 6576

QY 85 CCCAAA 90

Db 6575 GCCTAA 6570

RESULT 15

US-09-764-877-3978/C  
; Sequence 3978, Application US/09764877

; Patent No. US20020147140A1

; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.

; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies

; FILE REFERENCE: PC005

; CURRENT APPLICATION NUMBER: US/09/764,877

; CURRENT FILING DATE: 2001-01-17

; Prior application data removed - refer to PALM or file wrapper

; NUMBER OF SEQ ID NOS: 4031

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 3978

; LENGTH: 15071

; TYPE: DNA

; ORGANISM: Homo sapiens

US-09-764-877-3978

Query Match 29.1%; Score 27.6; DB 10; Length 15071;  
Best Local Similarity 63.6%; Pred. No. 3.7;  
Matches 42; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

QY 25 AGGAGAACAAAGACAGCCATTACCTGGAGAGACTGGCACTGATTTTACCCACAAG 84

Db 6635 AGAAAAAGAAAAAGAAAAAGATATGAAACAGACAGACACATCATGTGGCCCAAAAT 6576

QY 85 CCCAAA 90

Db 6575 GCCTAA 6570

Search completed: May 2, 2003, 15:02:30

Job time : 1219.22 secs





GenCore version 5.1.5  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 2, 2003, 12:54:45 ; Search time 689 Seconds

(Without alignments)  
8490.078 Million cell updates/sec

Title: US-09-719-554-3\_COPY\_7500\_7700

Perfect score: 201  
Sequence: 1 caactagtcgacacatca.....atttaccacccactgc 201

Scoring table: IDENTITY\_NTC  
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl: 1: gb\_ba: 2: gb\_htg: 3: gb\_in: 4: gb\_cm: 5: gb\_ov: 6: gb\_pat: 7: gb\_ph: 8: gb\_pl: 9: gb\_pr: 10: gb\_ro: 11: gb\_sts: 12: gb\_sy: 13: gb\_un: 14: gb\_vi: 15: em\_ba: 16: em\_fun: 17: em\_hum: 18: em\_in: 19: em\_mu: 20: em\_cm: 21: em\_or: 22: em\_ov: 23: em\_ph: 24: em\_pat: 25: em\_pl: 26: em\_ro: 27: em\_sts: 28: em\_un: 29: em\_vl: 30: em\_htg\_hum: 31: em\_htg\_inv: 32: em\_htg\_other: 33: em\_htg\_mus: 34: em\_htg\_pin: 35: em\_htg\_rod: 36: em\_htg\_mam: 37: em\_htg\_vrt: 38: em\_sy: 39: em\_higo\_hum: 40: em\_higo\_mus: 41: em\_higo\_other:

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	201	100.0	2055	6	AX007999 Sequence
2	201	100.0	2599	6	AX007978 Sequence
3	201	100.0	2930	6	AX355872 Sequence
4	201	100.0	2930	9	AF208161 Homo sapi
5	201	100.0	2946	6	AR177269 Sequence
6	201	100.0	10499	6	AX007980 Sequence
7	201	100.0	56093	6	AX329572 Sequence
8	201	100.0	56093	9	HSAC000064
9	201	100.0	149194	9	AC007566
10	199.4	99.2	7582	6	AX000966 Sequence
11	199.4	99.2	7582	6	AX027480
12	197.8	98.4	2006	6	AX000959 Sequence
13	197.8	98.4	2006	6	AX027473
14	197.8	98.4	2006	9	AF072503
15	185	92.0	2477	14	AF123882
16	185	92.0	158033	9	AC018926
17	185	92.0	187321	2	AC092510
18	184	91.5	155021	9	HS774G10
19	180.4	89.8	159851	9	AC036114
20	180.4	89.8	162356	6	AC018922
21	180.4	89.8	195344	2	AC025248
22	180.2	89.7	2575	6	AX000969 Sequence
23	180.2	89.7	46575	2	AC080036
24	180.2	89.7	114621	9	AC005187
25	180.2	89.7	170754	2	AC012403
26	180.2	89.7	177866	2	AL354698 Homo sapi
27	180.2	89.7	180954	9	AL160281 Human DNA
28	179	89.1	3372	6	AX0027478 Sequence
29	179	89.1	3372	9	AF072502
30	179	89.1	3372	2	AC009727
31	178.8	89.0	188607	2	AC008121
32	178.8	89.0	200926	2	AC008121 Homo sapi
33	178.6	88.9	1216	6	AX007984 Sequence
34	178.6	88.9	2784	6	AX007981 Sequence
35	178.6	88.9	103056	9	AC068447
36	178.6	88.9	135734	2	AC002346
37	178.6	88.9	161812	9	HS197017
38	178.6	88.9	177476	9	AC015983
39	178.6	88.9	195032	9	AL138706 Human DNA
40	178.6	88.9	205588	9	AC068400 Homo sapi
41	177.2	88.2	88823	9	AL713965 Human DNA
42	177	88.1	153444	2	AC040948
43	177	88.1	205035	9	CNS00005
44	172.6	85.9	73070	9	AL590143 Human DNA
45	172.6	85.9	160883	2	AL158814 Homo sapi

#### ALIGNMENTS

RESULT 1	LOCUS	AX007999	Sequence 22 from Patent WO967395.	2055 bp	DNA	linear	PAT 06-SEP-2000
AX007999	DEFINITION	Sequence 22 from Patent WO967395.					
AX007999	ACCESSION	AX007999.1	GI:9995696				
AX007999.1	VERSION						
AX007999.1	KEYWORDS						
AX007999.1	SOURCE						
AX007999.1	ORGANISM						

human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
REFERENCE  
1 (bases 1 to 2055)  
Perin,J.P., Rieger,F. and Alliel,P.M.  
Nucleic sequence and deduced protein sequence family with human  
endogenous retroviral motifs, and their uses

## JOURNAL

Patent: WO 9967395-A 22 29-DEC-1999;  
INST NAT SANTE RECH MED (FR); PERIN JEAN PIERRE (FR); RIEGER  
FRANCOIS (FR); ALIHEL PATRICK M (FR)

## FEATURES

Location/Qualifiers  
1..2055

## CDS

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
<1..87  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAC07392.1"  
/db\_xref="GI:9995697"  
/translation="PKTANLVADITSLAKYQVLTQGYTP"

BASE COUNT 576 a 574 c 376 g 529 t

## ORIGIN

Query Match 100.0%; Score 201; DB 6; Length 2055;  
Best Local Similarity 100.0%; Pred. No. 9.3e-52;  
Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

## QY

1 CAACTAGTTGACAGATCCTCCTTAGCCAAATATCAAGTTCTTAAACATTACA 60  
|||||

## DB

12 CAACTAGTTGACAGATCCTCCTTAGCCAAATATCAAGTTCTTAAACATTACA 71  
|||||

## QY

61 AGAACCTATCCCTGAGAGAGGAGAAAGACATATTCACCTGTGACATGTTAGT 120  
|||||

## DB

72 AGAACCTATCCCTGAGAGAGGAGAAAGACATATTCACCTGTGACATGTTAGT 131  
|||||

## QY

121 CAAGTCCTTCCCTTAATTCCTCCATCCCTAGATACATCTCTGGAGGAGCCATCCAGT 180  
|||||

## DB

132 CAAGTCCTTCCCTTAATTCCTCCATCCCTAGATACATCTCTGGAGGAGCCATCCAGT 191  
|||||

## QY

181 CATTATATCTACCCCAACTGC 201  
|||||

## DB

192 CATTATATCTACCCCAACTGC 212  
|||||

## RESULT 2

AX007978 AX007978 2599 bp DNA linear PAT 06-SEP-2000  
LOCUS Sequence 1 from Patent WO9967395.  
DEFINITION AX007978  
ACCESSION AX007978  
VERSION AX007978.1 GI:9995675  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
1 (bases 1 to 2599)  
AUTHORS Perin,J.P., Rieger,F. and Alihel,P.M.  
TITLE Nucleic sequence and deduced protein sequence family with human  
endogenous retroviral motifs, and their uses  
JOURNAL Patent: WO 9967395-A 1 29-DEC-1999;  
INST NAT SANTE RECH MED (FR); PERIN JEAN PIERRE (FR); RIEGER  
FRANCOIS (FR); ALIHEL PATRICK M (FR)

FEATURES Location/Qualifiers  
1..2599  
source /organism="Homo sapiens"  
/db\_xref="taxon:9606"

BASE COUNT 744 a 718 c 495 g 642 t

## ORIGIN

Query Match 100.0%; Score 201; DB 6; Length 2599;  
Best Local Similarity 100.0%; Pred. No. 9.5e-52;  
Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

## QY

1 CAACTAGTTGACAGATCCTCCTTAGCCAAATATCAAGTTCTTAAACATTACA 60  
|||||

## DB

550 CAACTAGTTGACAGATCCTCCTTAGCCAAATATCAAGTTCTTAAACATTACA 609  
|||||

## QY

61 AGAACCTATCCCTGAGAGAGGAGAAAGACATATTCACCTGTGACATGTTAGT 120  
|||||

## DB

610 AGAACCTATCCCTGAGAGAGGAGAAAGACATATTCACCTGTGACATGTTAGT 669  
|||||

QY 121 CAAGTCCTTCCCTTAATTCCTCCATCCCTAGATACATCTCTGGAGGAGCCATCCAGT 180  
|||||  
DB 670 CAAGTCCTTCCCTTAATTCCTCCATCCCTAGATACATCTCTGGAGGAGCCATCCAGT 729  
|||||

## QY

181 CATTATATCTACCCCAACTGC 201  
|||||

## DB

730 CATTATATCTACCCCAACTGC 750  
|||||

## RESULT 3

AX355872 AX355872 2930 bp DNA linear PAT 06-FEB-2002  
LOCUS Sequence 1 from Patent WO0204678.  
DEFINITION AX355872  
ACCESSION AX355872  
VERSION AX355872.1 GI:18620523  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
1  
AUTHORS Keith,J.C., McCoy,J.M. and M.L.S.  
TITLE Methods and compositions for diagnosing and treating preeclampsia  
and gestational trophoblast disorders  
JOURNAL Patent: WO 0204678-A 1 17-JAN-2002;  
GENETICS INSTITUTE, INC. (US)

FEATURES Location/Qualifiers  
1..2930  
source /organism="Homo sapiens"  
/db\_xref="taxon:9606"  
CDS 930..2546  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAD23000.1"  
/db\_xref="GI:18620524"

translation="MALPYHIFLFTVLPSFTLAPPQRCMTSSPYQELRMQRP  
GNIDAPYSRLSKGPTPTFAHMPKNCYSATLCMHANTHYTKMLIPSCGGLG  
IVCMVYFTQGMDSGGVODQAREKHKVEYISQLRVHGTSSPYKGLDLSKHETLRT  
HTRVLSLENTLTGLHEVSAQNPNTNCLPLNPRPVYSIPVPEOMNNSTELINTTSV  
IUGPLVSNLEITHHTSNLTCVRSNNTTYTNSGICWMPPTPIVRLPSIFPVCTSA  
YKLGSSSESKCFSLFVPEPTITTEODLYNIVISPRKRVPLIPVIGAVLQALG  
TGIGITFTSTQFYKLSDELNDMEVAVASLVTDQLNSLAVALVQNRALDLTAE  
NGTGLFTLEECYVYVNOGIVTEKVEKLRDRIRORAEELRNTGFMGLSQMMPILP  
FLGPLAAIILLFLFGPCIFNLVNFVSSRIEAVKIQMEPKMSKTKIYRRPLDRPASP  
RSDVMDIKGTPEPEISAAQPLRLPNSAGS"

BASE COUNT 842 a 800 c 571 g 717 t

Query Match 100.0%; Score 201; DB 6; Length 2930;  
Best Local Similarity 100.0%; Pred. No. 9.6e-52;  
Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAACTAGTTGACAGATCCTCCTTAGCCAAATATCAAGTTCTTAAACATTACA 60  
|||||

## DB

551 CAACTAGTTGACAGATCCTCCTTAGCCAAATATCAAGTTCTTAAACATTACA 610  
|||||

## QY

61 AGAACCTATCCCTGAGAGAGGAGAAAGACATATTCACCTGTGACATGTTAGT 120  
|||||

## DB

611 AGAACCTATCCCTGAGAGAGGAGAAAGACATATTCACCTGTGACATGTTAGT 670  
|||||

## QY

121 CAAGTCCTTCCCTTAATTCCTCCATCCCTAGATACATCTCTGGAGGAGCCATCCAGT 180  
|||||

## DB

671 CAAGTCCTTCCCTTAATTCCTCCATCCCTAGATACATCTCTGGAGGAGCCATCCAGT 730  
|||||

## QY

181 CATTATATCTACCCCAACTGC 201  
|||||

## DB

731 CATTATATCTACCCCAACTGC 751  
|||||

RESULT 4  
AF208161 AF208161 2930 bp mRNA linear PRI 22-FEB-2000  
LOCUS

D	b	611	AGGAACCTATCCCTCGAGAAGGAGAAAAAAGAAACTATTTCACCCTTGTCATGTATTTAGT	670
OY		121	C AAGTCCCTCCCTCCTTAATTTCCCATCCCTAGATACATCTCGGAGAGGCCCTAACCCAGT       	180
D	b	671	CAAGTCCCTCCCTCCTTAATTTCCCATCCCTAGATACATCTCGGAGAGGCCCTAACCCAGT	730
OY		181	CATTTTATCTACCCCCAACCTGC	201
D	b	731	CATTTTATCTACCCCCAACCTGC	751
R	E	S		
L	O	A	RI177269	2946 bp
D	F	I	Sequence 3 from patent US 6312921.	Linear
A	C	T	ARI177269	PAT 17-DEC-2001
V	E	R	ARI177269	
K	E	I	GI:17919624	
W	O	R		
S	O	R	Unknown.	
O	R	G	Unknown.	
R	E	F	Unclassified.	
A	H	T	1 (bases 1 to 2946)	
J	O	N	Jacobs,K., McCoy,J.M., Lavallie,E.R., Racine,L.A., Evans,C.,	
T	I	T	Morberg,D., Mi,S. and Treacy,M.	
J	O	N	Secreted proteins and polynucleotides encoding them	
F	E	A	Patent: US 6312921-A 3 06-NOV-2001;	
E	A	T	Location/Qualifiers	
S	O	R	1..2946	
B	A	S	/organism="unknown"	
C	O	U	858 a 801 c 570 g 717 t	
O	R	I		
Q	u	e		
r	y		Query Match	100.0%; Score 201; DB 6; Length 2946;
B	e	s	Best Local Similarity	100.0%; Pred. No.9,6e-52;
M	a	t	Matches 201; Conservative	0; Mismatches 0; Indels 0; Gaps 0;
O	Y			
1			C AACTTAGTTGGACAGACATCACCTCCTTAGCCAATATACAACAAGTCTTAAACATTAC	60
D	b	549	C AACTTAGTTGGACAGACATCACCTCCTTAGCCAATATACAACAAGTCTTAAACATTAC	608
O	Y	61	A GGAACCTATCCCTCGAGAAGGAGAAAAAAGAACTATTTCACCCTTGTCATGTATTTAGT	120
D	b	609	A GGAACCTATCCCTCGAGAAGGAGAAAAAAGAACTATTTCACCCTTGTCATGTATTTAGT	668
O	Y	121	C AATCCCTCCCTCCTTAATTTCCCATCCCTAGATACATCTCGGAGAGGCCCTAACCCAGT	180
D	b	669	C AATCCCTCCCTCCTTAATTTCCCATCCCTAGATACATCTCGGAGAGGCCCTAACCCAGT	728
O	Y	181	C ATTATCTACCCCCAACCTGC	201
D	b	729	C ATTATCTACCCCCAACCTGC	749
R	E	S		
L	O	A	AX007980	10499 bp
D	F	I	Sequence 3, from Patent WO9967395.	DNA
A	C	T	AX007980	Linear
V	E	R	AX007980.1 GI:9995677	PAT 06-SEP-2000
K	E	I		
S	O	R	human.	
O	R	G	Homo sapiens	
R	E	F	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
A	H	T	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
J	O	N	1 (bases 1 to 10499)	
F	E	A	Perin,J.P., Rieger,F. and Alliel,P.M.	
E	A	T	Nucleic sequence and deduced protein sequence family with human	
S	O	R	endogenous retroviral motifs, and their uses	
B	A	S	Patent: WO 9967395-A 3 29-DEC-1999;	
C	O	U	INST NAT SANTE RECH MED (FR); PERIN JEAN PIERRE (FR); RIEGER	
O	R	I	FRANCOIS (FR); ALLIEL PATRICK M (FR)	
F	E	A	Location/Qualifiers	
S	O	R	1..10499	

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
BASE COUNT 3048 a 2676 c 2280 g 2495 t  
ORIGIN

Query Match 100.0%; Score 201; DB 6; Length 10499;  
Best Local Similarity 100.0%; Pred. No. 1.1e-51;  
Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAAGTGTGACAGATCAGCTCCTAGCAATATACAAAGTCTTAAACATTACA 60  
|||||  
DB 7500 CAAGTGTGACAGATCAGCTCCTAGCAATATACAAAGTCTTAAACATTACA 7559  
QY 61 AGAAGCTATCCCTGAGAGAGGAGAAAGAACTATTCACCTTGTGACATGATTAGT 120  
|||||  
DB 7560 AGAAGCTATCCCTGAGAGAGGAGAAAGAACTATTCACCTTGTGACATGATTAGT 7619  
QY 121 CAAGTCCCTTCCCTTAATTCCTCCATCCCTAGATACATCTCTGGAGAGACCTACCCAGT 180  
|||||  
DB 7620 CAAGTCCCTTCCCTTAATTCCTCCATCCCTAGATACATCTCTGGAGAGACCTACCCAGT 7679  
QY 181 CATTATCTACCCCAACTGC 201  
|||||  
DB 7680 CATTATCTACCCCAACTGC 7700

RESULT 7  
AX329572 56093 bp DNA linear PAT 09-JAN-2002  
LOCUS AX329572  
DEFINITION Sequence 81 from Patent WO0194629.  
ACCESSION AX329572  
VERSION AX329572.1 GI:18102550  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE 1  
AUTHORS Young, P. E., Augustus, M., Carter, K. C., Ebner, R., Endress, G.,  
Horrigan, S., Soppet, D. R. and Weaver, Z.  
TITLE Cancer gene determination and therapeutic screening using signature  
gene sets  
JOURNAL Patent: WO 0194629-A 81 13-DEC-2001;  
Avalon Pharmaceuticals (US)  
FEATURES  
source 1. 56093  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"

BASE COUNT 16164 a 12346 c 10702 g 16881 t  
ORIGIN  
Query Match 100.0%; Score 201; DB 6; Length 56093;  
Best Local Similarity 100.0%; Pred. No. 1.2e-51;  
Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CAAGTGTGACAGATCAGCTCCTAGCAATATACAAAGTCTTAAACATTACA 60  
|||||  
DB 35500 CAAGTGTGACAGATCAGCTCCTAGCAATATACAAAGTCTTAAACATTACA 35559  
QY 61 AGAAGCTATCCCTGAGAGAGGAGAAAGAACTATTCACCTTGTGACATGATTAGT 120  
|||||  
DB 35560 AGAAGCTATCCCTGAGAGAGGAGAAAGAACTATTCACCTTGTGACATGATTAGT 35619  
QY 121 CAAGTCCCTTCCCTTAATTCCTCCATCCCTAGATACATCTCTGGAGAGACCTACCCAGT 180  
|||||  
DB 35620 CAAGTCCCTTCCCTTAATTCCTCCATCCCTAGATACATCTCTGGAGAGACCTACCCAGT 35679  
QY 181 CATTATCTACCCCAACTGC 201  
|||||  
DB 35680 CATTATCTACCCCAACTGC 35700

RESULT 8

HSAC000064 56093 bp DNA linear PRI 13-NOV-1996  
LOCUS HSAC000064  
DEFINITION Human BAC clone RG083M05 from 7q21-7q22, complete sequence.  
ACCESSION AC000064  
VERSION AC000064.1 GI:1669369  
KEYWORDS HTG.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens

REFERENCE 1  
AUTHORS Pauley, A.  
TITLE The sequence of H. sapiens BAC clone RG083M05  
JOURNAL Unpublished (1996)  
REFERENCE 2 (bases 1 to 56093)  
AUTHORS Waterston, R.  
TITLE Direct Submission  
JOURNAL Submitted (13-NOV-1996)  
COMMENT Genome Sequencing Center  
Department of Genetics, Washington University  
St. Louis, MO 63108, USA  
e-mail: sapiens@watson.wustl.edu

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping sections once, or longer because we provide a small overlap between neighboring submissions.

This sequence was finished as follows unless otherwise noted:  
all regions were double stranded or sequenced with an alternate chemistry; an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

SOURCE INFORMATION:  
This clone is from the first release of the human BAC library. The library contains cloned DNA from a human male fibroblast cell line 978SK. For references see: Shizuya et al., Proc. Natl. Acad. Sci. 89:8794-8797 (1992); Kim et al., Genomics 34:213-218 (1996).  
VECTOR: pBelo  
Selection: chloramphenicol

NEIGHBORING SEQUENCE INFORMATION:  
The orientation of this clone is unknown. Actual start of this clone is at base position 1 of H\_RG083M05; actual end is at 56093 of H\_RG083M05

This clone contains STS SWS1725.

FEATURES  
source 1. 56093  
Location/Qualifiers

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="7"  
/map="7q21-7q22"  
/clone.lib="CITB-978SK-B"  
complement(838..1131)  
/rpl\_family="ALU"

repeat\_region  
gene  
CDS  
/gene="WUSC:H\_RG083M05.1"  
join<1360..1503,4181..4370,4587..4774,6422..6556,9483..9547,11631..11773,11864..12021,13131..13296,14885..14988,16349..16546,16837..16971)  
/gene="WUSC:H\_RG083M05.1"  
/note="ATPase; strong similarity to peroxisome biosynthesis protein PAB1 (PID:91172019); coded for by human cDNA C04279 (NID:g1467530)"  
/codon\_start=1  
/protein\_id="BAB46346.1"  
/db\_xref="GI:1669371"  
/translation="KRLNIIQKTLLEAFSEAVMOPSVLLDLDLIGLPAVPEHEHSPDARCEIILCNVINKKIDCDINKFTDLDLGHVAKETGFAVRFYTVLVDAIRHSRL"



Db 35680 CATTIATCTACCCCACTGC 35700

```

|||||
RESULT 9
AC007566/c 149194 bp DNA linear PRI 01-MAR-2002
LOCUS Homo sapiens BAC clone CTB-10G5 from 7q21-7q22, complete sequence.
AC007566
AC007566.2 GI:11181861
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens.
Homo sapiens.
Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 149194)
Sulston, J.E. and Waterston, R.
Toward a complete human genome sequence
Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE
PUBMED
99063792
2 (bases 1 to 149194)
Du, Z.
The sequence of Homo sapiens BAC clone CTB-10G5
unpublished (2001)
3 (bases 1 to 149194)
Waterston, R.H.
Direct Submission
Submitted (15-MAY-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
4 (bases 1 to 149194)
Waterston, R.
Direct Submission
Submitted (02-OCT-2000) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
5 (bases 1 to 149194)
Waterston, R.H.
Direct Submission
Submitted (16-NOV-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
6 (bases 1 to 149194)
Waterston, R.H.
Direct Submission
Submitted (03-JAN-2002) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
7 (bases 1 to 149194)
Waterston, R.H.
Direct Submission
Submitted (06-FEB-2002) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
8 (bases 1 to 149194)
Waterston, R.
Direct Submission
Submitted (01-MAR-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Nov 16, 2000 this sequence version replaced g1:4835815.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: saplens@wustl.wustl.edu
----- Summary Statistics
Center Project name: H.KG010G05

```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

#### MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NIGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/CTB/CHR7>, send mailto:egreen@nhgri.nih.gov, or see <http://genome.wustl.edu/gsc>

#### SOURCE INFORMATION:

Clone CTB-10G5 is from the first release of the human BAC library CTB-978SK-B. The library contains cloned DNA from the male fibroblast cell line 978SK. See: Shizuya et al., Proc. Natl. Acad. Sci. USA 89:8794-7 (1992); U-J. Kim et al., Genomics 34:213-8 (1996). This clone is available from Research Genetics, Inc. (<http://www.resgen.com>).

VECTOR: pBelobAC11

Selection: chloramphenicol

#### NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP5-911H5, 200 base pair overlap. Actual start of this clone is at base position 195 of CTB-10G5; actual end is at base position 150532 of CTB-10G5.

FEATURES  
The clone CTB-10G5 contains the entire sequence of CTB-83M5.  
Location/Qualifiers

#### Source

```

1..149194
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/map="7q21-7q22"
/clone="CTB-10G5"
/clone_1lb="CTB-978SK-B"
1..1634
/rpt_family="L1"
2248..2387
/note="match to EST BG752883 (NID:g14063536)"
2248..2387
/note="match to EST AA149693 (NID:g1720635) zn99406.r1"
2248..2387
/note="match to EST AW579261 (NID:g7254310)"
2248..2387
/note="match to EST BG766882 (NID:g14077535)"
2248..2387
/note="match to EST BI160365 (NID:g14620366)"
2248..2374
/note="match to Homo sapiens EST BF758865 (NID:g12106765)"
2248..2287
/note="match to EST AV686676 (NID:g10286539)"
2253..2387
/note="match to EST AV123510 (NID:g10948226)"
2344..2387
/note="match to EST AV686676 (NID:g10286539)"
2696..3066
/rpt_family="MALR"
3108..3392
/rpt_family="Alu"
3540..3628
/note="match to EST AA425526 (NID:g2106267) zw48b03.r1"
3540..3628
/note="similar to Mus musculus EST BB253526 (NID:g8946272)"

```

misc_feature	3542..3632	/note="match to EST AV686676 (NID:g10288539) "
misc_feature	3542..3628	/note="match to EST AA149693 (NID:g1720635) zn99d06.r1"
misc_feature	3542..3628	/note="match to EST AU123510 (NID:g10948226) "
misc_feature	3542..3628	/note="match to EST AU123510 (NID:g10948226) "
misc_feature	3542..3628	/note="match to EST BG766882 (NID:g14077535) "
misc_feature	3542..3628	/note="match to EST B1160365 (NID:g14620366) "
misc_feature	3542..3609	/note="match to EST AW579261 (NID:g7254310) "
misc_feature	3562..3628	/note="match to EST BE272564 (NID:g9146913) "
misc_feature	3598..3628	/note="match to EST BE299708 (NID:g9183456) "
misc_feature	3714..3785	/note="match to EST BG260659 (NID:g12770475) "
misc_feature	3717..3785	/note="match to EST BG260659 (NID:g12770475) "
misc_feature	3719..3785	/note="similar to Mus musculus EST BE594936 (NID:g10678674) "
misc_feature	3719..3785	/note="match to EST AA149693 (NID:g1720635) zn99d06.r1"
misc_feature	3719..3785	/note="match to EST AA425526 (NID:g2106267) zw48b03.r1"
misc_feature	3719..3785	/note="match to EST AU123510 (NID:g10948226) "
misc_feature	3719..3785	/note="match to EST BE272564 (NID:g9146913) "
misc_feature	3719..3785	/note="match to EST BE299708 (NID:g9183456) "
misc_feature	3719..3785	/note="match to EST BG766882 (NID:g14077535) "
misc_feature	3719..3785	/note="match to EST B1160365 (NID:g14620366) "
misc_feature	3719..3785	/note="similar to Mus musculus EST BB253526 (NID:g8946272) "
misc_feature	3860..3862	/note="match to EST AV686676 (NID:g10288539) "
repeat_region	4730..4827	/rpt_family="L2"
repeat_region	4834..4925	/rpt_family="L1"
misc_feature	5715..5860	/note="match to EST AA425526 (NID:g2106267) zw48b03.r1"
misc_feature	5715..5860	/note="match to EST AU123510 (NID:g10948226) "
misc_feature	5715..5860	/note="match to EST BE272564 (NID:g9146913) "
misc_feature	5715..5860	/note="match to EST BE299708 (NID:g9183456) "
misc_feature	5715..5860	/note="match to EST BG260659 (NID:g12770475) "
misc_feature	5715..5860	/note="match to EST BG766882 (NID:g14077535) "
misc_feature	5715..5860	/note="match to EST B1160365 (NID:g14620366) "
misc_feature	5715..5860	/note="similar to Mus musculus EST BE594936

QY	121	CAAGTCCTCTCCCTCTAATTCCTCCATTCCTGATACATCCGGAAGAACCTACCACGT	180
Db	86384	CAAGTCCTCTCCCTCTAATTCCTCCATTCCTGATACATCTCGGGAAGAACCTACCACGT	86325
QY	181	CATTTATCTACCCCACTGC	201
Db	86324	CATTTATCTACCCCACTGC	86304
RESULT 10			
LOCUS	AX000966	7582 bp	DNA
DEFINITION	Sequence 11 from Patent WO9902696.	linear	PAT 10-MAR-2000
ACCESSION	AX000966		
VERSION	AX000966.1	GI:7241208	
KEYWORDS			
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS	unidentified.		
TITLE	unclassified.		
JOURNAL	unclassified.		
FEATURES			
source			
	1 (bases 1 to 7582)		
	Beesme, F. and Blond, J.		
	ENDOGENETIC RETROVIRAL SEQUENCES ASSOCIATED WITH AUTOIMMUNE		
	DISEASES OR WITH PREGNANCY DISORDERS		
	Patent: WO 9902696-A 11 21-JAN-1999;		
	BIO MERIEUX (FR); BESEME FREDERIC (FR)		
	Location/Qualifiers		
	1..7582		
	/organism="unidentified"		
	/db_xref="taxon:32644"		
BASE COUNT	2156 a 1876 c 1538 g 1796 t	216	others
ORIGIN			

Query Match	Score	DB	Length
Best Local Similarity 100.0%;	100.0%;	DB 9;	149194;
Matches 201; Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

  

QY	1	CAACTAGTTCGACAGACATCACTCTTAGCCAAATATCAACAAGTCTTAAACATTACA	60
DB <td>86504</td> <td>CAACTAGTTCGACAGACATCACTCTTAGCCAAATATCAACAAGTCTTAAACATTACA</td> <td>86445</td>	86504	CAACTAGTTCGACAGACATCACTCTTAGCCAAATATCAACAAGTCTTAAACATTACA	86445

  

QY	61	AGCAACTATCCCGAGAGAAAGGAAAGCAATCTCCACCCTTGACATGATTAAT	120
DB <td>86444</td> <th>AGCAACTATCCCGAGAGAAAGGAAAGCAATCTCCACCCTTGACATGATTAAT</th> <td>86385</td>	86444	AGCAACTATCCCGAGAGAAAGGAAAGCAATCTCCACCCTTGACATGATTAAT	86385

  

SOURCE	ORGANISM	HUMAN
REFERENCE	Human: Homo sapiens	
AUTHORS	Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Eumariola; Eutheria; Primates; Catarrhini; Homiidae; Homo.	
JOURNAL	1 (bases 1 to 7582)	
	Mallet, F., Voisset, C. and Paranhos, B. G.	
	Patent: FR 2788784-A 30 28-JUL-2000;	
	BIO MERIEUX (FR)	

  

FEATURES	location/Qualifiers
SOURCE	1..7582
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
BASE COUNT	2156 a 1876 c 1538 g 1796 t 216 others





Search completed: May 2, 2003, 15:26:45  
Job time : 759 secs



GenCore version 5.1.5  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 2, 2003, 12:49:35 ; Search time 159.333 seconds  
(without alignments)  
2840.910 Million cell updates/sec

Title: US-09-719-554-3\_COPY\_7500\_7700

Perfect score: 201

Sequence: 1 caactagtgtgcagacatca.....atttattctaccacacatgc 201

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq-10102:\*  
1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:\*  
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:\*  
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:\*  
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:\*  
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:\*  
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:\*  
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:\*  
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:\*  
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:\*  
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:\*  
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:\*  
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:\*  
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:\*  
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:\*  
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:\*  
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:\*  
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:\*  
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:\*  
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:\*  
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:\*  
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:\*  
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:\*  
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:\*  
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	201	100.0	2055	21	ABN97948
2	201	100.0	2599	21	ABN97927
3	201	100.0	2930	24	AA24195
4	201	100.0	2946	20	AA27526
5	201	100.0	2946	21	AA59468
6	201	100.0	10489	21	ABN97929
7	201	100.0	56093	24	AB161744
8	199.4	99.2	7582	20	AA25665
9	199.4	99.2	7582	21	AA59215

10	197.8	98.4	2006	20	AA25658	Human endogenous r
11	197.8	98.4	2006	21	AA59208	3' pol gene and 5'
12	185	92.0	1148	23	AA584200	DNA encoding novel
13	180.2	89.7	2575	20	AA25668	Human endogenous r
14	179	89.1	3372	20	AA25663	Human endogenous r
15	179	89.1	3372	21	AA59213	Partial pol gene a
16	179	89.1	3831	23	AA57127	DNA encoding novel
17	179	89.1	5154	23	AA567609	DNA encoding novel
18	179	89.1	7466	23	AA568626	DNA encoding novel
19	179	89.1	8279	23	AA576474	DNA encoding novel
20	178.6	88.9	1216	21	ABN97933	Human retroviral s
21	178.6	88.9	2784	21	ABN97930	Human retroviral s
22	170.8	85.0	1740	21	ABN97973	Human retroviral s
23	153	76.1	8294	23	AA584209	DNA encoding novel
24	150.8	75.0	2585	20	AA25667	DNA encoding novel
25	146.6	72.9	542	19	AA15134	Human endogenous r
26	146.6	72.9	542	19	AAV09124	Human adult testes
27	109.8	54.6	40668	24	ABQ88150	Nucleotide sequenc
28	108.8	54.1	2186	23	AA576465	Human osteoblast d
29	100.8	50.1	758	19	AA43322	DNA encoding novel
30	100.8	50.1	758	20	AA29705	Multiple sclerosis
31	99.4	49.5	4535	23	AA576205	Clone LB16 from MS
32	90.2	44.9	611	23	AA571721	DNA encoding novel
33	90.2	44.9	1218	23	AA571717	DNA encoding novel
34	90.2	44.9	1218	23	AA584193	DNA encoding novel
35	90.2	44.9	1219	23	AA573812	DNA encoding novel
36	88.4	44.0	1963	23	AA591202	DNA encoding novel
37	50	24.9	387	23	AA584202	DNA encoding novel
38	50	24.9	1948	20	AA25659	Human endogenous r
39	50	24.9	1948	21	AA59209	5' non coding, 3'
40	50	24.9	2781	22	AA55630	Nucleotide sequenc
41	50	24.9	2782	20	AA25661	Human endogenous r
42	50	24.9	2782	21	AA59211	5' non coding, 3'
43	50	24.9	2782	22	AAH20069	HERV-W envelope pr
44	46.4	23.1	200	22	ABA50726	Human breast cell
45	46.4	23.1	200	22	ABA68694	Human foetal liver

## ALIGNMENTS

RESULT 1  
ABN97948  
ID ABN97948 standard; DNA: 2055 BP.  
AC ABN97948:  
XX  
XX  
DT 01-AUG-2002 (first entry)  
DE Human retroviral HERV-7g env coding sequence.  
KW Autoimmune disease; HERV-7g; chromosome 7q; immunotherapy;  
KW multiple sclerosis; ds.  
XX  
XX Human retrovirus.  
OS  
XX  
XX PN WO967395-A1.  
XX PD 29-DEC-1999.  
XX PF 23-JUN-1999; 99MO-FR01513.  
XX PR 23-JUN-1998; 98FR-0007920.  
XX  
XX (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.  
XX Alliel PM, Perlin J, Rieger F;  
XX WPI; 2000-160587/14.  
XX New nucleic acid sequences of human endogenous retrovirus, HERV-7g,  
XX used for diagnosis, treatment and prevention of autoimmune and  
XX neurological diseases -

XX Claim 3; Page 142-145; 225pp; French.  
PS  
XX  
CC The present invention relates to new nucleic acid sequences of human  
CC endogenous retrovirus, HERV-7g, which is located on chromosome 7g.  
CC Regulatory elements associated with HERV-7g may alter expression of other  
CC genes (even remote genes) on the same chromosome, inducing immunological  
CC and/or neurological changes (which may be pathological or protective/  
CC curative). HERV-7g peptides can be used to improve efficiency of the  
CC immune response, e.g. in immunotherapy. HERV-7g peptides and their coding  
CC sequences can be used in immunogenic or vaccinating compositions, for  
CC protection against autoimmune diseases, particularly multiple sclerosis.  
CC The peptides may also be used (by sequence comparison) to detect/identify  
CC endogenous retroviruses that are abnormally expressed in cancer,  
CC neuropathologies or other autoimmune diseases. The present sequence was  
CC used to illustrate the invention.  
XX  
SQ Sequence 2055 BP; 576 A; 574 C; 376 G; 529 T; 0 other;  
Query Match 100.0%; Score 201; DB 21; Length 2055;  
Best Local Similarity 100.0%; Pred. No. 1.3e-54;  
Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 CAACCTAGTGGAGACATCACCCTCTAGCCAAATATCAAGTCTTAAACATTACA 60  
DB 12 CAACCTAGTGGAGACATCACCCTCTAGCCAAATATCAAGTCTTAAACATTACA 71  
QY 61 AGGAACCTATCCCTGAGAGAGGAAAGAACTATTCACCCCTGTGACATGTAATAGT 120  
DB 72 AGGAACCTATCCCTGAGAGAGGAAAGAACTATTCACCCCTGTGACATGTAATAGT 131  
QY 121 CAAGTCCCTCCCTCTAATTTCCCATCCTAGATACATCTGTGGAGAGACCTTACCAGT 180  
DB 132 CAAGTCCCTCCCTCTAATTTCCCATCCTAGATACATCTGTGGAGAGACCTTACCAGT 191  
QY 181 CATTTATCTACCCCACTGC 201  
DB 192 CATTTATCTACCCCACTGC 212  
RESULT 2  
ABN97927  
ID ABN97927 standard; DNA; 2599 BP.  
XX  
AC ABN97927;  
XX  
DF 01-AUG-2002 (first entry)  
XX  
DE Human retroviral sequence HERV 7 env.  
XX  
XX Autoimmune disease; HERV-7g; chromosome 7g; immunotherapy;  
KM multiple sclerosis; ds.  
XX  
OS Human retrovirus.  
XX  
FN WO9967395-A1.  
XX  
PD 29-DEC-1999.  
XX  
PE 23-JUN-1999; 99WO-FR01513.  
XX  
PR 23-JUN-1998; 98FR-0007920.  
XX  
PA (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.  
XX  
PI Alliel PM, Berlin J, Rieger F;  
XX  
DR WPI; 2000-160587/14.  
XX  
PT New nucleic acid sequences of human endogenous retrovirus, HERV-7g,  
PT used for diagnosis, treatment and prevention of autoimmune and  
PT neurological diseases  
XX

PS Claim 1; Page 128-129; 225pp; French.  
XX  
XX  
CC The present invention relates to new nucleic acid sequences of human  
CC endogenous retrovirus, HERV-7g, which is located on chromosome 7g.  
CC Regulatory elements associated with HERV-7g may alter expression of other  
CC genes (even remote genes) on the same chromosome, inducing immunological  
CC and/or neurological changes (which may be pathological or protective/  
CC curative). HERV-7g peptides can be used to improve efficiency of the  
CC immune response, e.g. in immunotherapy. HERV-7g peptides and their coding  
CC sequences can be used in immunogenic or vaccinating compositions, for  
CC protection against autoimmune diseases, particularly multiple sclerosis.  
CC The peptides may also be used (by sequence comparison) to detect/identify  
CC endogenous retroviruses that are abnormally expressed in cancer,  
CC neuropathologies or other autoimmune diseases. The present sequence was  
CC used to illustrate the invention.  
XX  
SQ Sequence 2599 BP; 744 A; 718 C; 495 G; 642 T; 0 other;  
Query Match 100.0%; Score 201; DB 21; Length 2599;  
Best Local Similarity 100.0%; Pred. No. 1.4e-54;  
Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 CAACCTAGTGGAGACATCACCCTCTAGCCAAATATCAAGTCTTAAACATTACA 60  
DB 550 CAACCTAGTGGAGACATCACCCTCTAGCCAAATATCAAGTCTTAAACATTACA 609  
QY 61 AGGAACCTATCCCTGAGAGAGGAAAGAACTATTCACCCCTGTGACATGTAATAGT 120  
DB 610 AGGAACCTATCCCTGAGAGAGGAAAGAACTATTCACCCCTGTGACATGTAATAGT 669  
QY 121 CAAGTCCCTCCCTCTAATTTCCCATCCTAGATACATCTGTGGAGAGACCTTACCAGT 180  
DB 670 CAAGTCCCTCCCTCTAATTTCCCATCCTAGATACATCTGTGGAGAGACCTTACCAGT 729  
QY 181 CATTTATCTACCCCACTGC 201  
DB 730 CATTTATCTACCCCACTGC 750  
RESULT 3  
AAD24195  
ID AAD24195 standard; CDNA; 2930 BP.  
XX  
AC AAD24195;  
XX  
DF 07-MAY-2002 (first entry)  
XX  
DE Human syncytin cDNA.  
XX  
XX Human; syncytin; pre-eclampsia; gestational trophoblast disorder;  
KM Chorioncarcinoma; hydatiform mole; placental site tumour; abortion;  
KW envelope gene; human endogenous defective retrovirus; HERV-W; ss.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT CDS 930..2546  
FT /tag=a  
FT /product="Syncytin"  
XX  
FN WO200204678-A2.  
XX  
PD 17-JAN-2002.  
XX  
PE 09-JUL-2001; 2001WO-US21719.  
XX  
PR 07-JUL-2000; 2000US-216657P.  
XX  
PA (GENY ) GENETICS INST INC.  
XX  
PI Keith JC, McCoy JM, M1 S;  
XX  
DR WPI; 2002-171727/22.  
XX

DR P-PSDB: AAE14540.  
XX Identifying a compound for treating a subject with or at risk of  
PT developing pre-eclampsia, comprises determining whether the expression  
PT or activity of syncytin in the cell is modulated in the presence of a  
PT test compound  
XX  
PS Disclosure; Page 39-42; 43pp; English.  
XX  
CC The invention relates to identifying compounds which are modulators  
CC of syncytin expression. The syncytin modulators are useful in diagnosis  
CC and treatment of pre-eclampsia and gestational trophoblast disorders (e.g.  
CC choriocarcinoma, hydatiform mole, placental site tumour and missed/  
CC incomplete abortion). Syncytin is a human gene derived from the  
CC envelope gene of human endogenous defective retrovirus, HERV-W. The  
CC present invention is based partly on the discovery that syncytin  
CC expression is dramatically reduced in pre-eclampsia, and is also  
CC mis-localised to the apical syncytiotrophoblast membrane. The present  
CC sequence is human syncytin cDNA.  
XX  
SQ Sequence 2930 BP; 842 A; 800 C; 571 G; 717 T; 0 other;  
Query Match 100.0%; Score 201; DB 24; Length 2930;  
Best Local Similarity 100.0%; Pred. No. 1.5e-54;  
Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 CAACTAGTTCGAGACATCACCCTCTTACCAATATACACAAGTCTTAAACATTACA 60  
DB 551 CAACCTAGTTCGAGACATCACCCTCTTACCAATATACACAAGTCTTAAACATTACA 610  
OY 61 AGAAGCCTATCCCTGAGAGAGGAAAGAACTATTCACCCCTTGATGATTTAGT 120  
DB 611 AGAAGCCTATCCCTGAGAGAGGAAAGAACTATTCACCCCTTGATGATTTAGT 670  
OY 121 CAAGTCCTTCCTCTAATTCCTCCATCCCTAGATATCTGGGAGAGACCTACCAGT 180  
DB 671 CAAGTCCTTCCTCTAATTCCTCCATCCCTAGATATCTGGGAGAGACCTACCAGT 730  
OY 181 CATTTATCTACCCCACTGC 201  
DB 731 CATTTATCTACCCCACTGC 751  
RESULT 4  
AAZ7526  
ID AAZ7526 standard; cDNA; 2946 BP.  
XX  
AC AAZ7526;  
XX  
DT 10-AUG-1999 (first entry)  
XX  
DE Human secreted protein AJ172\_2 cDNA.  
XX  
KW Secreted protein; testes; brain; blood; placenta; human; murine; thymus;  
KW bone marrow; treatment; prevention; cytokine; immune; vaccine;  
KW cell proliferation; cell differentiation; suppressor; tumour inhibition;  
KW haematopoiesis regulator; activin; inhibin; chemotactic; chemokine;  
KW haemostatic; thrombolytic; receptor; ligand; anti-inflammatory; tumour;  
KW cadherin; tumour invasion suppressor; gene therapy; tissue growth; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO9926972-A1.  
XX  
PD 03-JUN-1999.  
XX  
XX 17-NOV-1998; 98WO-US24614.  
XX  
XX 20-OCT-1998; 98US-0175928.  
XX 21-NOV-1997; 97US-0976110.  
XX 18-MAY-1998; 98US-0080478.  
XX  
PA (GENY ) GENETICS INST INC.

XX  
PI Collins-Racie LA, Evans C, Jacobs K, Lavallie ER;  
PI McCoy JM, Meberg D, Treacy M;  
XX WPI: 1999-357813/30.  
DR P-PSDB: AAY08622.  
XX  
PT New polynucleotides encoding secreted proteins  
XX  
PS Claim 13a; Page 100-101; 142pp; English.  
XX  
CC This invention describes novel human secreted proteins encoded by  
CC polynucleotides isolated from human adult testes, adult brain, adult  
CC blood or adult placenta, or murine adult bone marrow or thymus cDNA  
CC libraries. The products of the invention are predicted to have biological  
CC activities which would make them suitable for treating, preventing or  
CC ameliorating medical conditions in humans and animals, although no  
CC supporting data is given. Suggested activities include nutritional  
CC activity, cytokine and cell proliferation/differentiation activity,  
CC immune stimulating (e.g. as vaccines) or suppressing activity,  
CC haematopoiesis regulating activity, tissue growth activity,  
CC activin/inhibin activity, chemotactic/chemokine activity, haemostatic  
CC and thrombolytic activity, receptor/ligand activity, anti-inflammatory  
CC activity, cadherin/tumour invasion suppressor activity, and tumour  
CC inhibition activity. The polynucleotides are also stated to be useful  
CC for gene therapy.  
XX  
SQ Sequence 2946 BP; 858 A; 801 C; 570 G; 717 T; 0 other;  
Query Match 100.0%; Score 201; DB 20; Length 2946;  
Best Local Similarity 100.0%; Pred. No. 1.5e-54;  
Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 CAACTAGTTCGAGACATCACCCTCTTACCAATATACACAAGTCTTAAACATTACA 60  
DB 549 CAACCTAGTTCGAGACATCACCCTCTTACCAATATACACAAGTCTTAAACATTACA 608  
OY 61 AGAAGCCTATCCCTGAGAGAGGAAAGAACTATTCACCCCTTGATGATTTAGT 120  
DB 609 AGAAGCCTATCCCTGAGAGAGGAAAGAACTATTCACCCCTTGATGATTTAGT 668  
OY 121 CAAGTCCTTCCTCTAATTCCTCCATCCCTAGATATCTGGGAGAGACCTACCAGT 180  
DB 669 CAAGTCCTTCCTCTAATTCCTCCATCCCTAGATATCTGGGAGAGACCTACCAGT 728  
OY 181 CATTTATCTACCCCACTGC 201  
DB 729 CATTTATCTACCCCACTGC 749  
RESULT 5  
AAZ59468  
ID AAZ59468 standard; cDNA; 2946 BP.  
XX  
AC AAZ59468;  
XX  
DT 11-APR-2000 (first entry)  
XX  
DE Human secreted protein AJ172\_2 polynucleotide sequence.  
XX  
KW Human; secreted protein; disease diagnosis; pre-eclampsia; cancer;  
KW placental pathology; metastasis inhibition; nutritional activity;  
KW immune stimulator; haematopoiesis regulator; tissue growth;  
KW tumour inhibitor; anti-inflammatory; clone AJ172\_2; ATCC\_98115;  
KW gene therapy; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO9960020-A1.  
XX  
XX 25-NOV-1999.  
XX  
XX 17-MAY-1999; 99WO-US10915.

XX 18-MAY-1998; 98US-0080478.  
PR 20-OCT-1998; 98US-0175928.  
XX  
XX (GEMV ) GENETICS INST INC.  
PI Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;  
PI Werberg D, Ml S, Treacy M;  
DR WPI: 2000-116311/10.  
DR P-PSDB: AAY67313.  
XX  
PT New polynucleotides encoding secreted cDNA libraries, used to develop  
PT products for the diagnosis and treatment of neoplastic disease  
XX  
PS Claim 14; Page 107-108; 149pp; English.  
XX  
CC This is the human secreted protein AJ172\_2 nucleotide sequence, obtained  
CC from a human adult testes cDNA library. The invention relates to secreted  
CC human and murine proteins. The polynucleotides and proteins are predicted  
CC to have biological activities which would make them suitable for  
CC treating, preventing or ameliorating medical conditions in humans and  
CC animals. Detection of the levels of the proteins can be used for the  
CC diagnosis of e.g. pre-eclampsia, placental pathology or cancer. Agents  
CC which modulate the expression or function of the proteins may be used for  
CC treating a neoplastic disease and inhibiting metastasis. Other suggested  
CC activities include nutritional activity (e.g. in feeds), cytokine and  
CC cell proliferation/differentiation activity, immune stimulating  
CC (e.g. as vaccines) or suppressing activity, haematopoiesis regulating  
CC activity, tissue growth activity, activin/inhibin activity,  
CC chemotactic/chemokinetic activity, haemostatic and thrombolytic activity,  
CC receptor/ligand activity, anti-inflammatory activity, cadherin/tumour  
CC invasion suppressor activity, and tumour inhibition activity. The  
CC polynucleotide sequences are also stated to be useful for gene therapy.  
XX  
SQ Sequence 2946 BP; 858 A; 801 C; 570 G; 717 T; 0 other;  
Query Match 100.0%; Score 201; DB 21; Length 2946;  
Best Local Similarity 100.0%; Pred. No. 1.5e-54;  
Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 CAACCTAGTGGACATCCCTCTAGCCAAATATCAACAGTCTTAAACATTACA 60  
DB 549 CAACCTAGTGGACATCCCTCTAGCCAAATATCAACAGTCTTAAACATTACA 608  
QY 61 AGAACCTATCCCTGAGAGAGGAAAGAACTATTCCACCCTTGTGACATGCTATTAGT 120  
DB 609 AGAACCTATCCCTGAGAGAGGAAAGAACTATTCCACCCTTGTGACATGCTATTAGT 668  
QY 121 CAAGTCCCTTCCCTTAATTCCTCCATCCCTAGATACATCTGGGAAAGACCTACCCAGT 180  
DB 669 CAAGTCCCTTCCCTTAATTCCTCCATCCCTAGATACATCTGGGAAAGACCTACCCAGT 728  
QY 181 CATTATATCAACCCCAACTGC 201  
DB 729 CATTATATCAACCCCAACTGC 749  
RESULT 6  
ABN97929  
ID ABN97929 standard; DNA: 10499 BP.  
XX  
XX ABN97929;  
XX  
DT 01-AUG-2002 (first entry)  
XX  
XX Human retroviral sequence HERV-7q.  
XX  
XX Autimmune disease; HERV-7q; chromosome 7q; immunotherapy;  
KW multiple sclerosis; ds.  
XX  
XX Human retrovirus.  
XX

PN WO967395-A1.  
XX  
PD 29-DEC-1999.  
XX  
XX 23-JUN-1999; 99WO-FR01513.  
PF  
XX 23-JUN-1998; 98FR-0007920.  
PR  
XX (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.  
PA  
PI Allel PM, Perlin J, Rieger F;  
PI WPI: 2000-160587/14.  
DR  
XX  
XX  
PT New nucleic acid sequences of human endogenous retrovirus, HERV-7q,  
PT used for diagnosis, treatment and prevention of autoimmune and  
PT neurological diseases  
XX  
PS Claim 3; Fig 1; 225pp; French.  
XX  
XX  
CC The present invention relates to new nucleic acid sequences of human  
CC endogenous retrovirus, HERV-7q, which is located on chromosome 7q.  
CC Regulatory elements associated with HERV-7q may alter expression of other  
CC genes (even remote genes) on the same chromosome, inducing immunological  
CC and/or neurological changes (which may be pathological or protective/  
CC curative). HERV-7q peptides can be used to improve efficiency of the  
CC immune response, e.g. in immunotherapy. HERV-7q peptides and their coding  
CC sequences can be used in immunogenic or vaccinating compositions, for  
CC protection against autoimmune diseases, particularly multiple sclerosis.  
CC The peptides may also be used (by sequence comparison) to detect/identify  
CC endogenous retroviruses that are abnormally expressed in cancer,  
CC neuropathologies or other autoimmune diseases. The present sequence was  
CC used to illustrate the invention.  
XX  
SQ Sequence 10499 BP; 3048 A; 2676 C; 2280 G; 2495 T; 0 other;  
Query Match 100.0%; Score 201; DB 21; Length 10499;  
Best Local Similarity 100.0%; Pred. No. 2.3e-54;  
Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 CAACCTAGTGGACATCCCTCTAGCCAAATATCAACAGTCTTAAACATTACA 60  
DB 7500 CAACCTAGTGGACATCCCTCTAGCCAAATATCAACAGTCTTAAACATTACA 7559  
QY 61 AGAACCTATCCCTGAGAGAGGAAAGAACTATTCCACCCTTGTGACATGCTATTAGT 120  
DB 7560 AGAACCTATCCCTGAGAGAGGAAAGAACTATTCCACCCTTGTGACATGCTATTAGT 7619  
QY 121 CAAGTCCCTTCCCTTAATTCCTCCATCCCTAGATACATCTGGGAAAGACCTACCCAGT 180  
DB 7620 CAAGTCCCTTCCCTTAATTCCTCCATCCCTAGATACATCTGGGAAAGACCTACCCAGT 7679  
QY 181 CATTATATCAACCCCAACTGC 201  
DB 7680 CATTATATCAACCCCAACTGC 7700  
RESULT 7  
ABL61744  
ID ABL61744 standard; DNA: 56093 BP.  
XX  
XX ABL61744;  
XX  
DT 15-MAY-2002 (first entry)  
XX  
XX Colon adenocarcinoma related gene sequence SEQ ID NO:81.  
XX  
XX Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;  
KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;  
KW cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;  
KW gene; ds.  
XX  
XX Homo sapiens.  
OS

XX WO200194629-A2.  
XX 13-DEC-2001.  
XX 30-MAY-2001; 2001WO-US10838.  
XX 05-JUN-2000; 2000US-209473P.  
XX 05-JUN-2000; 2000US-209531P.  
XX 18-SEP-2000; 2000US-233133P.  
XX 18-SEP-2000; 2000US-233617P.  
XX 20-SEP-2000; 2000US-234009P.  
XX 20-SEP-2000; 2000US-234034P.  
XX 20-SEP-2000; 2000US-234052P.  
XX 22-SEP-2000; 2000US-234509P.  
XX 22-SEP-2000; 2000US-234567P.  
XX 25-SEP-2000; 2000US-234923P.  
XX 25-SEP-2000; 2000US-234923P.  
XX 25-SEP-2000; 2000US-235077P.  
XX 25-SEP-2000; 2000US-235082P.  
XX 25-SEP-2000; 2000US-235134P.  
XX 25-SEP-2000; 2000US-235280P.  
XX 26-SEP-2000; 2000US-235637P.  
XX 26-SEP-2000; 2000US-235638P.  
XX 27-SEP-2000; 2000US-235711P.  
XX 27-SEP-2000; 2000US-235720P.  
XX 27-SEP-2000; 2000US-235840P.  
XX 27-SEP-2000; 2000US-235863P.  
XX 28-SEP-2000; 2000US-236028P.  
XX 28-SEP-2000; 2000US-236032P.  
XX 28-SEP-2000; 2000US-236033P.  
XX 28-SEP-2000; 2000US-236034P.  
XX 28-SEP-2000; 2000US-236109P.  
XX 28-SEP-2000; 2000US-236111P.  
XX 29-SEP-2000; 2000US-236842P.  
XX 29-SEP-2000; 2000US-236891P.  
XX 02-OCT-2000; 2000US-237172P.  
XX 02-OCT-2000; 2000US-237173P.  
XX 02-OCT-2000; 2000US-237278P.  
XX 02-OCT-2000; 2000US-237294P.  
XX 02-OCT-2000; 2000US-237295P.  
XX 02-OCT-2000; 2000US-237316P.  
XX 03-OCT-2000; 2000US-237425P.  
XX 03-OCT-2000; 2000US-237598P.  
XX 03-OCT-2000; 2000US-237604P.  
XX 03-OCT-2000; 2000US-237606P.  
XX 03-OCT-2000; 2000US-237608P.  
XX 01-NOV-2000; 2000US-244867P.  
XX 01-NOV-2000; 2000US-245084P.  
XX (AVAL-) AVALON PHARM.  
XX Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;  
PI Soppet DR, Weaver Z;  
XX WPI: 2002-188264/24.  
XX  
XX WPI: 2002-188264/24.  
XX  
XX Claim 1; SEQ ID 81; 44pp; English.  
XX  
XX The present invention describes a method (M1) for screening for an  
CC anti-neoplastic agent. The method involves exposing cells to a chemical  
CC agent to be tested for anti-neoplastic activity, determining a change in  
CC expression of at least one gene (I) of a signature gene set, where (I)  
CC comprises a sequence (S) selected from 8447 sequences (given in AB16164  
CC to AB170110), or is at least 95% identical to (S), where a change in  
CC expression is indicative of anti-neoplastic activity. (I) has cytostatic  
CC activity and can be used in gene therapy. M1 can be used for screening  
CC an anti-neoplastic agent, and can be used for producing a product which  
CC is the data collected with respect to the anti-neoplastic agent as a

CC result of M1, and the data is sufficient to convey the chemical  
CC structure and/or properties of the agent. M1 can be used in the  
CC treatment of cancer such as colon, breast, stomach, lung, thyroid,  
CC oesophageal, ovarian, kidney, prostate or pancreatic cancer,  
CC adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,  
CC infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine  
CC carcinoma, papillary carcinoma and Wilms' tumour.  
XX  
SQ Sequence 56093 BP; 16164 A; 12346 C; 10702 G; 16881 T; 0 other;  
Query Match 100.0%; Score 201; DB 24; Length 56093;  
Best Local Similarity 100.0%; Pred. No. 4e-54;  
Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 CAAGTACCTGAGACATACCTCTTACGCAATATACACAGTCTTAAACATTACA 60  
DB 35500 CAAGTACCTGAGACATACCTCTTACGCAATATACACAGTCTTAAACATTACA 35559  
OY 61 AGGAACCTATCCCTGAGAGAGGAGAAAGAACTATTCACCCCTGTGACATGTATTAGT 120  
DB 35560 AGGAACCTATCCCTGAGAGAGGAGAAAGAACTATTCACCCCTGTGACATGTATTAGT 35619  
OY 121 CAAGTCCCTTCCCTTAATTCCTCCATCCCTAGATACATCTGGAAGACCTACCCAGT 180  
DB 35620 CAAGTCCCTTCCCTTAATTCCTCCATCCCTAGATACATCTGGAAGACCTACCCAGT 35679  
OY 181 CATTTATCTACCCCACTGC 201  
DB 35680 CATTTATCTACCCCACTGC 35700  
RESULT 8  
AAX25665  
ID AAX25665 standard; cDNA to mRNA; 7582 BP.  
XX  
XX AAX25665;  
AC  
XX 21-MAY-1999 (first entry)  
DT  
XX  
DE Complete human endogenous retrovirus W genome.  
XX  
XX Clone; human endogenous retrovirus; genome; autoimmune disease;  
XX multiple sclerosis; rheumatoid polyarthritis; insulin-dependent diabetes;  
XX disseminated lupus erythematosus; pregnancy; chromosomal marker; ss.  
XX  
XX Human endogenous retrovirus.  
OS  
XX  
XX WO9902696-A1.  
XX  
XX 21-JAN-1999.  
PD  
XX  
XX 06-JUL-1998; 98WO-FR01442.  
PF  
XX  
XX 07-JUL-1997; 97FR-0008815.  
PR  
XX  
XX (INMR ) BIO MERIEUX.  
PA  
XX  
XX Beseme F, Blond JL, Bouton O, Mallet F, Mandrand B;  
PI  
XX  
XX WPI: 1999-120897/10.  
DR  
XX  
XX New nucleic acid sequences from human endogenous retrovirus-W -  
PT expressed exclusively in placenta and useful in diagnosis and  
PT therapy of autoimmune disease, and abnormal or failed pregnancy  
XX  
XX Claim 1; Page 71-74; 106pp; French.  
XX  
XX This sequence represents the complete sequence of the human endogenous  
CC retrovirus (HERV) W genome. The nucleic acids, their fragments or  
CC peptides encoded by them are markers of autoimmune disease (e.g. multiple  
CC sclerosis, rheumatoid polyarthritis, disseminated lupus erythematosus,  
CC insulin-dependent diabetes and related pathologies) and of abnormal or  
CC unsuccessful pregnancy and can be used as chromosomal markers for





PA (INNR ) BIO MERIEUX.  
XX Beseme F, Blond JL, Bouton O, Mallet F, Mandrand B;  
XX WPI: 1999-120897/10.  
XX  
XX New nucleic acid sequences from human endogenous retrovirus-W -  
PT expressed exclusively in placenta and useful in diagnosis and  
PT therapy of autoimmune disease, and abnormal or failed pregnancy  
XX  
XX Claim 1; Page 54-56; 106pp; French.  
XX  
XX This sequence represents clone c1.P122 of the human endogenous retrovirus  
CC (HERV) W genome. The nucleic acids, their fragments or peptides encoded  
CC by them are markers of autoimmune disease (e.g. multiple sclerosis,  
CC rheumatoid polyarthritis, disseminated lupus erythematosus, insulin-  
CC dependent diabetes and related pathologies) and of abnormal or  
CC unsuccessful pregnancy and can be used as chromosomal markers for  
CC susceptibility to these conditions, or proximity markers of genes  
CC associated with this susceptibility.  
XX  
SQ Sequence 2006 BP; 625 A; 532 C; 373 G; 475 T; 1 other:  
  
Query Match 98.4%; Score 197.8; DB 20; Length 2006;  
Best Local Similarity 99.0%; Pred. No. 1.4e-53;  
Matches 199; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
OY 1 CAACCTAGTGGAGACATCACCCTCCTTAGCCAAATATCAACAAGTCTTAAACATTACA 60  
DB 1042 CAACCTAGTGGAGACATCACCCTCCTTAGCCAAATATCAACAAGTCTTAAACATTACA 1101  
OY 61 AGAACATATCCCTGAGAGAGAGAAAGAACTATTCACCCCTTGACATGTAATTAGT 120  
DB 1102 AGAACATATCCCTGAGAGAGAGAAAGAACTATTCACCCCTTGACATGTAATTAGT 1161  
OY 121 CAAGTCCCTCCCTCTAATTCCTCCATGATACATCCTGGAGAGACCTTACCAGT 180  
DB 1162 CAAGTCCCTCCCTCTAATTCCTCCATGATACATCCTGGAGAGACCTTACCAGT 1221  
OY 181 CATTTATCTACCCCACTGC 201  
DB 1222 CATTTATCTACCCCACTGC 1242  
  
RESULT 11  
AAS59208  
ID AAS59208 standard; DNA; 2006 BP.  
XX  
AC AAS59208;  
XX  
DT 07-NOV-2000 (first entry)  
XX  
DE 3' pol gene and 5' env gene sequences of HERV-W from human genome.  
XX  
KW Autoimmune disease; retrovirus; human endogenous retrovirus W; HERV-W;  
KW gag gene; pregnancy; multiple sclerosis; T cell proliferation; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200043521-A2.  
XX  
PD 27-JUL-2000.  
XX  
PF 21-JAN-2000; 2000WO-FR00144.  
XX  
PR 21-JAN-1999; 99FR-0000888.  
XX  
PA (INNR ) BIO MERIEUX.  
XX  
PI Paranhos-Baccala G, Mallet F, Voisset C;  
XX  
DR WPI: 2000-499229/44.  
XX

PT New nucleic acid from human endogenous retrovirus, useful e.g. for  
PT diagnosis of autoimmune disease and complications of pregnancy,  
PT contains at least part of the gag gene -  
XX  
XX Disclosure; Page 44-45; 53pp; French.  
XX  
XX The present sequence represents an endogenetic retroviral nucleic acid  
CC fragment, which is associated with an autoimmune disease, and is  
CC integrated into the human genome. The fragment is originally derived  
CC from a novel retrovirus, human endogenous retrovirus W (HERV-W). The  
CC HERV-W retrovirus is associated with autoimmune disease, failure of  
CC pregnancy or disorders of pregnancy. The nucleic acid fragment, or  
CC proteins derived from it, are useful for diagnosis of autoimmune  
CC disease (specifically multiple sclerosis) and for monitoring pregnancy.  
CC The nucleic acid fragments may also be used for in situ labelling of  
CC isolated chromosomes, while the transcription product can be used to  
CC study or monitor T cell proliferation in vitro.  
XX  
SQ Sequence 2006 BP; 625 A; 532 C; 373 G; 475 T; 1 other:  
  
Query Match 98.4%; Score 197.8; DB 21; Length 2006;  
Best Local Similarity 99.0%; Pred. No. 1.4e-53;  
Matches 199; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
  
OY 1 CAACCTAGTGGAGACATCACCCTCCTTAGCCAAATATCAACAAGTCTTAAACATTACA 60  
DB 1042 CAACCTAGTGGAGACATCACCCTCCTTAGCCAAATATCAACAAGTCTTAAACATTACA 1101  
OY 61 AGAACATATCCCTGAGAGAGAGAAAGAACTATTCACCCCTTGACATGTAATTAGT 120  
DB 1102 AGAACATATCCCTGAGAGAGAGAAAGAACTATTCACCCCTTGACATGTAATTAGT 1161  
OY 121 CAAGTCCCTCCCTCTAATTCCTCCATGATACATCCTGGAGAGACCTTACCAGT 180  
DB 1162 CAAGTCCCTCCCTCTAATTCCTCCATGATACATCCTGGAGAGACCTTACCAGT 1221  
OY 181 CATTTATCTACCCCACTGC 201  
DB 1222 CATTTATCTACCCCACTGC 1242  
  
RESULT 12  
AAS84200  
ID AAS84200 standard; CDNA; 1148 BP.  
XX  
AC AAS84200;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE DNA encoding novel human diagnostic protein #20004.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US08631.  
XX  
PR 31-MAR-2000; 2000US-0540217.  
XX  
PR 23-AUG-2000; 2000US-0649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;  
XX  
DR WPI: 2001-639362/73.  
XX  
DR P-PSDB; ABG20013.  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in

PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -

Claim 1; SEQ ID No 20004; 103pp; English.

The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome and gene mapping, and in recombinant production of (II). The polynucleotides are also used in diagnostics as expressed sequence tags for identifying expressed genes. (I) is useful in gene therapy techniques to restore normal activity of (II) or to treat disease states involving (II). (II) is useful for generating antibodies against it, detecting or quantitating a polypeptide in tissue, as molecular weight markers and as a food supplement. (I) and its binding partners are useful in medical imaging of sites expressing (II). (I) and (II) are useful for treating disorders involving aberrant protein expression or biological activity. The polypeptide and polynucleotide sequences have applications in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits to assess biodiversity and to produce other types of data at a products dependent on DNA and amino acid sequences. AA564197-AA594174 represent novel human diagnostic coding sequences of the invention.

Note: The sequence data for this patent did not appear in the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published/pct\\_sequences](http://wipo.int/pub/published/pct_sequences).

Sequence 1148 BP; 368 A; 286 C; 242 G; 252 T; 0 other;

Query Match	92.08	Score 185	DB 23	Length 1148
Best Local Similarity	95.08	Pred. No. 1.5e-49		
Matches 191; Conservative	0	Mismatches 10	Indels 0	Gaps 0

QY	1	CAAGTTAGTGCACAGATCAACCTCTTAGCCAA	ATATCAACAGAGTCTTTAAACATTAC	60
Db	797	CAAGTTAGTGCACAGATCAACCTCTTAGCCAA <td>ATATCAACAGAGTCTTTAAACATTAC</td> <td>856</td>	ATATCAACAGAGTCTTTAAACATTAC	856
QY	61	AGGAACCTATCCCGAGAAAGAGGAAAGAAAC <td>ATTCACACCTCTGTGACATGGTATTAAG</td> <td>120</td>	ATTCACACCTCTGTGACATGGTATTAAG	120
Db	857	AGGAGCCTGTGCCGAGAGAGGAGAAAGAAAT <td>TTTCCACCTCGTGTCTACGTGATTAAG</td> <td>916</td>	TTTCCACCTCGTGTCTACGTGATTAAG	916
QY	121	CAAGTCCCTTCCCTCTTAATTCGCCATCCATG <td>AGATACATCTCTGGGAAAGAGACCTACCAAG</td> <td>180</td>	AGATACATCTCTGGGAAAGAGACCTACCAAG	180
Db	917	CAAGTCCCTTCCCTCTTAATTCGCCATCCATG <td>AGATACATCTCTGGGAGAGACCTACCAAG</td> <td>976</td>	AGATACATCTCTGGGAGAGACCTACCAAG	976
QY	181	CAATTTATCTACCCCAACTGC	201	
Db	977	CAATTTATCTATCCCAACTGC	997	

RESULT 13  
AAAX25668  
ID AAX25668 standard: cDNA to mRNA; 2575 Bf.

AC AAX25668;

DT 21-MAY-1999 (first entry)

Human endogenous retrovirus W clone cl.7321.

**kw** Clone; human endogenous retrovirus; genotype; autoimmune disease;

disseminated lupus erythematosus; pregnancy; chromosomal marker; ss.

OS Human endogenous retrovirus.

PN WO9902696-A1.

PD 21-JAN-1999.

PF 06-JUL-1998; 98WO-FR01442.

PR 07-JUL-1997; 97FR-0008815.

PA (INMR) BIO MERIEUX.

PI Beseme F, Blond JL, Bouton O, Mallet F, Mandrand B;

DR WPI; 1999-120897/10.

PT New nucleic acid sequences from human endogenous retrovirus-W -

PT therapy of autoimmune disease, and abnormal or failed pregnancy

PS Claim 1; Page 80-83; 106pp; French.

This sequence represents clone c1.7A21 of the human endogenous retrovirus (HERV) W genome. The nucleic acids, their fragments or peptides encoded by them are markers of autoimmune diseases (e.g. multiple sclerosis, rheumatoid polyarthritis, disseminated lupus erythematosus, insulin-dependent diabetes and related pathologies) and of abnormal or unsuccessful pregnancy and can be used as chromosomal markers for susceptibility to these conditions, or proximity markers of genes associated with this susceptibility.

Sequence 2575 BP; 817 A; 608 C; 546 G; 602 T; 2 other;

Query Match	89.78;	Score 180.2;	DB 20;	Length 2575;
-------------	--------	--------------	--------	--------------

```
Matches 188; Conservative 0; Mismatches 13; Indels 0; Gaps 0;
```

QY	1	CAATTGTTGGCAGACATCACTCCCTTAGCCAAATATCAACAAGTTCTTAAACCAATTACA	60
Db	2285	CAACTTGTTGGCAGACATCACTCCCTTAGCCAAATATCAACAAGTTCTTAAACCAATTACA	2344
QY	61	AGGAACTATCCCTGGAGAAAGGGAAAGAAACATATCCACCCTTGACATGATTTGT	120
Db	2345	AGGAACTATCCCTGGAGAAAGGGAAAGGAAACATATCCACCCTTGACATGATTTGT	2404
QY	121	CAAGTCCCTTCCTCTATATTCGCCATCCCTAGATACATCTCTGGAGAGACCTACCAGT	180
Db	2405	CAAGTCCCTTCCTCTATATTCCTATTCCTAGTATATATCTCTGGAGAGACCTACCAGT	2466
QY	181	CAATTTATCTACCCCAACTGC	201
Db	2465	CAATTTATCTACCCCAACTGC	2485

RESULT 14  
AAAX25663  
ID AAAX25663 standard; cDNA to mRNA; 3372 BP.

AAAX25663;

DT 21-MAY-1999 (first entry)

DE Human endogenous retrovirus W clone cl.P15T.

KW Clone; human endogenous retrovirus; genome; autoimmune disease; ...

KW disseminated lupus erythematosus; pregnancy; chromosomal marker; ss.

OS Human endogenous retrovirus.

PN WO9902696-A1.

PD 21-JAN-1999.

06-JUL-1998; 98WO-FR01442.

PR 07-JUL-1997; 97FR-0008815

PA ( INMR ) BIO MERIEUX .

PI Beseme F, Blond JL, Bouton O, Mallet F, Mandrand B;

XX DR WPI; 1999-120897/10.  
XX PT New nucleic acid sequences from human endogenous retrovirus-W -  
XX PT expressed exclusively in placenta and useful in diagnosis and  
XX PT therapy of autoimmune disease, and abnormal or failed pregnancy  
XX PS Claim 1; Page 64-67; 106pp; French.  
XX This sequence represents clone cl.P15F of the human endogenous retrovirus  
CC (HERV) W genome. The nucleic acids, their fragments or peptides encoded  
CC by them are markers of autoimmune disease (e.g. multiple sclerosis,  
CC rheumatoid polyarthritis, disseminated lupus erythematosus, insulin-  
CC dependent diabetes and related pathologies) and of abnormal or  
CC unsuccessful pregnancy and can be used as chromosomal markers for  
CC susceptibility to these conditions, or proximity markers of genes  
CC associated with this susceptibility.  
XX SQ Sequence 3372 BP; 1047 A; 835 C; 711 G; 779 T; 0 other;  
XX  
XX Query Match 89.1%; Score 179; DB 20; Length 3372;  
Best Local Similarity 96.5%; Pred. No. 1.9e-47;  
Matches 194; Conservative 0; Mismatches 5; Indels 2; Gaps 1;  
QY 1 CAACTAGTGCAGACATCACTCTTAGCCAAATATCAACAAGTCTTAAACATTACA 60  
DB 2431 CAACTAGTGCAGACATCACTCTTAGCCAAATATCAACAAGTCTTAAACATTACA 2490  
QY 61 AGGAACCTATCCCTGAGAGAAGGAGAAAGAACTATTCACCCCTTGATGATGATTAGT 120  
DB 2491 AGGAACCTATCCCTGAGAGAAGG--AAAAGAAATATTCACCCCAAGTATGATTAGT 2548  
QY 121 CAACTCCCTTCCCTCTAATTCCTCCATCCCTAGATACATCCTGGGAAGACCTACCCAGT 180  
DB 2549 CAACTCCCTTCCCTCTAATTCCTCCATCCCTAGATACATCCTGGGAAGACCTACCCAGT 2608  
QY 181 CATTATATCTACCCCAACTGC 201  
DB 2609 CATTATATCTACCCCAACTGC 2629  
RESULT 15  
ID AAA59213  
XX AAA59213 standard; DNA; 3372 BP.  
XX AC AAA59213;  
XX DT 07-NOV-2000 (first entry)  
XX DE Partial pol gene and U3-R region sequences of HERV-W from human genome.  
XX KM Autoimmune disease; retrovirus; human endogenous retrovirus W; HERV-W;  
XX KM gag gene; pregnancy; multiple sclerosis; T cell proliferation; ss.  
XX OS Homo sapiens.  
XX PN WO200043521-A2.  
XX PD 27-JUL-2000.  
XX PF 21-JAN-2000; 2000WO-FR00144.  
XX PR 21-JAN-1999; 99FR-0000888.  
XX PA (INMR) BIO MERIEUX.  
XX PI Paranhos-Baccala G, Mallet F, Voisset C;  
XX DR WPI; 2000-499229/44.  
XX PT New nucleic acid from human endogenous retrovirus, useful e.g. for  
PT diagnosis of autoimmune disease and complications of pregnancy,  
PT contains at least part of the gag gene

XX PS Disclosure; Page 47-48; 53pp; French.  
XX CC The present sequence represents an endogenous retroviral nucleic acid  
CC fragment, which is associated with an autoimmune disease, and is  
CC integrated into the human genome. The fragment is originally derived  
CC from a novel retrovirus, human endogenous retrovirus W (HERV-W). The  
CC HERV-W retrovirus is associated with autoimmune disease, failure of  
CC pregnancy or disorders of pregnancy. The nucleic acid fragment, or  
CC proteins derived from it, are useful for diagnosis of autoimmune  
CC disease (specifically multiple sclerosis) and for monitoring pregnancy.  
CC The nucleic acid fragments may also be used for in situ labelling of  
CC isolated chromosomes, while the transcription product can be used to  
CC study or monitor T cell proliferation in vitro.  
XX SQ Sequence 3372 BP; 1047 A; 835 C; 711 G; 779 T; 0 other;  
XX  
XX Query Match 89.1%; Score 179; DB 21; Length 3372;  
Best Local Similarity 96.5%; Pred. No. 1.9e-47;  
Matches 194; Conservative 0; Mismatches 5; Indels 2; Gaps 1;  
QY 1 CAACTAGTGCAGACATCACTCTTAGCCAAATATCAACAAGTCTTAAACATTACA 60  
DB 2431 CAACTAGTGCAGACATCACTCTTAGCCAAATATCAACAAGTCTTAAACATTACA 2490  
QY 61 AGGAACCTATCCCTGAGAGAAGGAGAAAGAACTATTCACCCCTTGATGATGATTAGT 120  
DB 2491 AGGAACCTATCCCTGAGAGAAGG--AAAAGAAATATTCACCCCAAGTATGATTAGT 2548  
QY 121 CAACTCCCTTCCCTCTAATTCCTCCATCCCTAGATACATCCTGGGAAGACCTACCCAGT 180  
DB 2549 CAACTCCCTTCCCTCTAATTCCTCCATCCCTAGATACATCCTGGGAAGACCTACCCAGT 2608  
QY 181 CATTATATCTACCCCAACTGC 201  
DB 2609 CATTATATCTACCCCAACTGC 2629

Search completed: May 2, 2003, 14:51:19  
Job time : 184.333 secs



GenCore version 5.1.5  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 2, 2003, 14:19:30 ; Search time 112 Seconds  
(without alignments)  
2927.422 Million cell updates/sec

Title: US-09-719-554-3\_COPY\_7500\_7700

Perfect score: 201  
Sequence: 1 caactagttgcagacatca.....atttattaccccaactgc 201

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 809774376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estnu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hlc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hlc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vrt:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_rtd:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	201	100.0	362	13	BI049622-CM3-GN029
2	201	100.0	503	12	BG000981-110-GN024
3	177	88.1	578	9	AA613238
4	153	76.1	414	14	AA613238 no19f04.s
5	144.8	72.0	171	12	HI2439 yj11n10.r1
6	101.8	50.6	605	17	BG013341 CM4-GN036
					AO240663 CIT-HSP-2

Result No.	Score	Query Match	Length	ID	Description
7	97.6	48.6	496	10	BE208551
8	97	48.3	618	17	B59957
9	95.2	47.4	516	17	AO428067
10	91	45.3	536	10	AM074844
11	89	44.3	672	17	B68542
12	86.4	43.0	138	12	BG013345
13	84.6	42.1	578	12	B54046
14	82.6	41.1	563	17	AO427133
15	80	39.8	477	17	AO309705
16	78.8	39.2	330	14	N76991
17	70.6	35.1	483	17	AO609732
18	69.2	34.4	408	13	BI034898
19	67.4	33.5	347	14	B0024998
20	62	30.8	403	17	AO632527
21	61.4	30.5	426	13	BI033854
22	60	29.9	723	13	BI772113
23	57	28.4	448	17	BI7253
24	55.8	27.8	574	17	AO634644
25	55.8	27.8	637	17	AO378897
26	48.6	24.2	422	17	AO228090
27	48	23.9	363	17	AO651656
28	48	23.9	390	17	AO059215
29	47	23.4	751	13	BM014497
30	46	22.9	668	12	BE730884
31	43.8	21.8	451	17	AO184628
32	40.4	20.1	290	14	B0022458
33	39.2	19.5	576	17	B52708
34	39	19.4	561	17	AO487402
35	39	19.4	614	17	B75173
36	38	18.9	939	17	CNS00CNG
37	36.2	18.0	700	17	AO262910
38	35	17.4	609	17	A2796232
39	35	17.4	710	13	BG914193
40	35	17.4	1735	12	BE790606
41	34.8	17.3	546	10	BE684788
42	34.6	17.2	513	9	AA498640
43	34.6	17.2	629	17	AG125475
44	34.4	17.1	438	17	B99018
45	34	16.9	945	17	CNS03NCM

## ALIGNMENTS

RESULT 1  
LOCUS BI049622 362 bp mRNA linear EST 15-JUN-2001  
DEFINITION CM3-GN0297-020101-591-g08 GN0297 Homo sapiens cDNA, mRNA sequence.  
ACCESSION BI049622  
VERSION BI049622.1 GI:14457152  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 362)  
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Britones,M.R., Nagai,M.A., da Silva,W.U., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bata,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J., and Simpson,A.J.  
TITLE Shotgun sequencing of the human transcriptome with ORF expressed  
JOURNAL Sequence tags  
MEDLINE Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
COMMENT Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001



High quality sequence stop: 442.

FEATURES  
Location/Qualifiers

1..578

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:1101151"

/tissue="pneumocystoma"

/lab\_host="SOLR (kanamycin resistant)"

/note="Vector: Bluescript SK-; Site: EcoRI; Site 2: XhoI"

/note="Cloned unidirectionally. Primer: Oligo dT."

/note="Pneumocystoma, 5' adaptor sequence: 5' GATTGGCAGCAG 3' 3' adaptor sequence: 5' CTCGAGTTTCTTTTCTTTT 3' "

Average insert size: 1.3 Kb."

BASE COUNT  
128 a 116 c 160 g 174 t

ORIGIN

Query Match

Best Local Similarity

Matches 186; Conservative

0; Mismatches 15; Indels 0; Gaps 0;

Score 177; DB 9; Length 578;

Pred. No. 4.8e-44;

1 CACTTACTGACACATCACCCTCTTACCAATATACAGTTCTTAAACATTACA 60

237 CAACCTAGTTCAGATATACCTCTTACCAATATACAGTTCTTAAACATTACA 178

61 AGAAGCTATCCCTGAGAGAGGAGAAAGAACTATTCACCTGTGACATGTTAGT 120

177 AGAAGCTATCCCTGAGAGAGGAGAAAGAACTATTCACCTGTGACATGTTAGT 118

121 CAAGTCCCTTCCCTTAATTCCTTACCTGATACATCTCTGGAAGAGCCCTACCT 180

117 CAAGTCCCTTCCCTTCTGATCCCTGATACATCTCTGGAAGAGCCCTACCT 58

181 CATTTATCTACCCCACTGC 201

57 CATTTATCTACCCCACTGC 37

DB

RESULT 4

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

Insert size: 678

High quality sequence stops: 257

Source: IMAGE Consortium, LNL

This clone is available royalty-free through LNL; contact the

IMAGE Consortium (info@image.lnl.gov) for further information.

Insert Length: 678 Std Error: 0.00

Seq primer: M13RPI

High quality sequence stop: 257.

FEATURES

Location/Qualifiers

1..414

/organism="Homo sapiens"

/db\_xref="GDB:560252"

/db\_xref="taxon:9606"

/clone="IMAGE:148483"

/clone="Soares Placenta NB2HP"

/sex="Female"

/dev\_stage="placenta obtained at birth (full term)"

/lab\_host="DH10B (ampicillin resistant)"

/note="Organ: placenta; Vector: pT73D (Pharmacia) with a

modified polylinker; Site: Not I; Site 2: Eco RI; 1st

strand cDNA was primed with a Not I - oligo(dT) primer [5'

ACTGGAAGATTCGCGCCGAGAGATTTTCTTTTCTTTT 3'],

double-stranded cDNA was ligated to Eco RI adaptors

(Pharmacia), digested with Not I and cloned into the Not I

and Eco RI sites of the modified pT73 vector. Library

went through one round of normalization. Library

constructed by Bento Soares and M.Fatima Bonaldo.

1 others

BASE COUNT

119 a 107 c 109 g 109 t

ORIGIN

Query Match

Best Local Similarity

Matches 164; Conservative

0; Mismatches 0; Indels 1; Gaps 1;

Score 153; DB 14; Length 414;

Pred. No. 1.1e-36;

1 CACTTACTGACACATCACCCTCTTACCAATATACAGTTCTTAAACATTACA 60

149 CACTTACTGACACATCACCCTCTTACCAATATACAGTTCTTAAACATTACA 208

61 AGAAGCTATCCCTGAGAGAGGAGAAAGAACTATTCACCTGTGACATGTTAGT 120

209 AGAAGCTATCCCTGAGAGAGGAGAAAGAACTATTCACCTGTGACATGTTAGT 268

121 CAAGTCCCTTCCCTTAATTCCTTACCTGATACATCTCTGGAAGAGCCCTACCT 164

269 CAAGTCCCTTCCCTTCTGATCCCTGATACATCTCTGGAAGAGCCCTACCT 313

DB

RESULT 5

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

COMMENT

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

Insert size: 678

High quality sequence stops: 257

Source: IMAGE Consortium, LNL

This clone is available royalty-free through LNL; contact the

IMAGE Consortium (info@image.lnl.gov) for further information.

Insert Length: 678 Std Error: 0.00

Seq primer: puc 18 forward

High quality sequence start: 13

## High quality sequence stop: 17..

FEATURES  
source location/Qualifiers  
1..171

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="GN0361"  
/dev\_stage="Adult"  
/note="Organ: Placenta; normal; Vector: puc18; Site\_1: SmaI; Site\_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent Application No. 196,776 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT  
54 a 48 c 26 g 43 t

Query Match  
Best Local Similarity 98.6%; Score 144.8; DB 12; Length 171;  
Matches 146; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CAACCTAGTGCAGACATCACTCCCTTACGCCAATATCAACAAGTTCTTAAACATTACA 60  
11 GCACCTAGTGCAGACATCACTCCCTTACGCCAATATCAACAAGTTCTTAAACATTACA 70  
QY 61 AGAACCTATCCCTGAGAGAGGAGAAAGACTATTCACCCCTTGACATGGTATTAGT 120  
71 AGAACCTATCCCTGAGAGAGGAGAAAGACTATTCACCCCTTGACATGGTATTAGT 130  
DB 121 CAAGTCCTTCCCTCTATATCCCATCC 148  
131 CAAGTCCTTCCCTCTATATCCCATCC 158

RESULT 6  
LOCUS AQ240663 605 bp DNA linear GSS 30-SEP-1998  
DEFINITION CIT-HSP-2385F10.TF.1 CIT-HSP Homo sapiens genomic clone 2385F10,  
ACCESSION AQ240663  
VERSION AQ240663.1 GI:3672861  
KEYWORDS GSS.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
REFERENCE 1 (bases 1 to 605)  
AUTHORS Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K., Berry,K., Granger,D., Sub,E., White,C., Shizuya,H., Simon,M. and Venter,J.C.  
TITLE Use of a random human BAC End Sequence Database for Sequence-Ready Map Building  
JOURNAL Unpublished (1998)  
COMMENT Other\_GSSs: CIT-HSP-2385F10.TR.1  
Contact: Mark Adams  
Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: mdadams@igf.org  
Clones are available from Research Genetics (info@resgen.com). BAC  
clone search page: [http://www.tigr.org/cdb/humgen/bac\\_end\\_search/bac\\_end\\_search.html](http://www.tigr.org/cdb/humgen/bac_end_search/bac_end_search.html).  
Seq primer: M13-21  
Class: BAC ends.

FEATURES  
source location/Qualifiers  
1..605

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="2385F10"  
/clone\_lib="CIT-HSP"  
/sex="Male"

/cell\_type="Sperm"  
/note="Vector: pBeloBac11; Site\_1: HindIII; Site\_2: HindIII"  
BASE COUNT  
189 a 157 c 119 g 140 t

Query Match  
Best Local Similarity 69.2%; Score 101.8; DB 17; Length 605;  
Matches 139; Conservative 0; Mismatches 62; Indels 0; Gaps 0;

QY 1 CAACCTAGTGCAGACATCACTCCCTTACGCCAATATCAACAAGTTCTTAAACATTACA 60  
293 CAACCTAGTGCAGACATCACTCCCTTACGCCAATATCAACAAGTTCTTAAACATTACC 352  
DB 61 AGAACCTATCCCTGAGAGAGGAGAAAGACTATTCACCCCTTGACATGGTATTAGT 120  
353 TGAAGATGTCACAGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 412  
QY 121 CAAGTCCTTCCCTCTATATCCCATCCATCCATGATATCTCGGAGAGACCTAACCT 180  
413 CAATCTCTCCCTCTATATCCCATCCATGATATCTCGGAGAGACCTAACCTAACCT 472  
QY 181 CATTATCTACCCCACTGC 201  
DB 473 AATCTCTCTATCTCCCACTGC 493

RESULT 7  
LOCUS BE208551/c 496 bp mRNA linear EST 27-JUN-2000  
DEFINITION ba09a03.y1 NIH\_MGC\_7 Homo sapiens cDNA clone IMAGE:2823820 5' similar to SW:POL.SMSAV P03359 POL POLYPROTEIN [CONTAINS: REVERSE TRANSCRIPTASE]; mRNA sequence.  
ACCESSION BE208551  
VERSION BE208551.1 GI:8751949  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
REFERENCE 1 (bases 1 to 496)  
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>  
TITLE NIH-MGC National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)  
COMMENT Other\_ESTs: ba09a03.x1  
Contact: Robert Strausberg, Ph.D.  
Email: c9apbs-r@mail.nih.gov  
cDNA Library Preparation: Ling Hong/Rubin Laboratory cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov/image/html/resources.shtml>

Possible reversed clone: similarity on wrong strand  
Seq primer: -40RP from Glibco  
High quality sequence stop: 493.  
FEATURES  
source location/Qualifiers  
1..496

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_image="2823820"  
/clone\_lib="NIH\_MGC\_7"  
/clone\_type="small cell carcinoma"  
/cell\_line="MGC3"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: Lung; Vector: pORF7; Site\_1: XhoI; Site\_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(C). Size-selected >500bp for average insert size 1.8kb. Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit





DB 64 TACATACGTGGAGAGACCCCTACCCAGTCACTTATCTACCTGATGTC 111

RESULT 10  
AM074844/c 536 bp mRNA linear EST 20-OCT-2000  
LOCUS AM074844  
DEFINITION XA08C04.X1 Soares NFL T.GBC.S1 Homo sapiens cDNA clone  
IMAGE:2567718.3 similar to SW:P01\_SMSAV P03359 POLYPROTEIN  
[CONTAINS: REVERSE TRANSCRIPTASE, mRNA sequence.]

ACCESSION  
AM074844.1 GI:6029842  
EST.

KEYWORDS

SOURCE

ORGANISM

human.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1 (bases 1 to 536)

NCI-CCGAP http://www.ncbi.nlm.nih.gov/ncicgap.

AUTHORS

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

JOURNAL

Unpublished (1997)

CONTACT: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

This clone is available royalty-free through LNL; contact the

IMAGE Consortium (info@image.llnl.gov) for further information.

Insert length: 831 Std Error: 0.00

Seq primer: -400p from gibco

High quality sequence stop: 459.

FEATURES

source

1.536

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="IMAGE:2567718"

/clone\_11b="Soares\_NFL\_T.GBC.S1"

/lab\_host="DH10B"

/note="Organ: pooled; vector: pT73D-Pac (Pharmacia) with

a modified polylinker; Site\_1: Not I; Site\_2: Eco RI;

Equal amounts of plasmid DNA from three normalized

libraries (fetal lung fibroblasts, testis NHT, and B-cell

NCI-CGAP (GCB1) were mixed, and ss circles were made in

vitro. Following HAP purification, this DNA was used as

tracer in a subtractive hybridization reaction. The driver

was PCR-amplified cDNAs from pools of 5,000 clones made

from the same 3 libraries. The pools consisted of

I.M.A.G.E. clones 297460-302087, 682632-687239,

726408-728711, and 729036-731399. Subtraction by Bento

Soares and M. Fatima Bonaldi.

BASE COUNT

141 a 105 c 136 g 154 t

ORIGIN

Query Match

Best Local Similarity 66.7%; Score 91; DB 10; Length 536;

Matches 130; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

QY 1 CAACCTAGTGGAGACATCACTCTTACCCAAATCAACAAGTCTTAAACATTACA 60

DB 405 CAATTTGGTCAATATATATCTTTGGCTAAATATCAATGAAACCTTAAACCCACC 346

QY 61 AGAAGCTATCCCTGAGAGAGAGAGAAAGAACTATCCACCCCTTGACATGTTAGT 120

DB 345 TGAAGATGTCAACAGAGAGAGAGAGAGAGTGTTCACACGAGAGATCTAGTGTGG 286

QY 121 CAAGTCCCTCCCTCTAATTCCTCCCTAGTATCAATCTGGAGAGAGAGAGAGAGAG 180

DB 285 CAAATCTCTCCCTCTACCTCCCTATGATTTCTTGTGGAGAGAGAGAGAGAGAG 226

QY 181 CATTATCTACCC 195

DB 225 AATCTCTCTGCTC 211

RESULT 11

B68542

LOCUS B68542 672 bp DNA linear GSS 21-JUN-1998  
CIT-HSP-205314.TF CIT-HSP Homo sapiens genomic clone 205314, DNA  
DEFINITION sequence.

ACCESSION  
B68542  
VERSION  
B68542.1 GI:2667252  
KEYWORDS  
GSS.

SOURCE

human.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1 (bases 1 to 672)

Adams,M.D., Rounsley,S.D., Field,C.E., Bass,S., Linher,K., Golden

,K., Berry,K., Granger,D., Sun,E., Wible,C., Shizuya,H., Simon,M.

and Venter,J.C.

Use of a random BAC End Sequence Database for Sequence-Ready Map

Building

JOURNAL

Unpublished (1997)

CONTACT: Mark Adams

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: mdamad@tigr.org

Clones are available from Research Genetics (info@resgen.com). BAC

end search page:

http://www.tigr.org/tdb/human/gen/bac\_end\_search/bac\_end\_search.html

Seq primer: M13-21

Class: BAC ends.

FEATURES

source

1.672

/organism="Homo sapiens"

/db\_xref="GDB:7057735"

/db\_xref="taxon:9606"

/clone="205314"

/clone\_11b="CIT-HSP"

/sex="Male"

/cell\_type="Sperm"

/note="Vector: pBelobAC11; Site\_1: HindIII; Site\_2:

HindIII"

BASE COUNT

206 a 161 c 137 g 168 t

ORIGIN

Query Match

Best Local Similarity 65.2%; Score 89; DB 17; Length 672;

Matches 131; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

QY 1 CAACCTAGTGGAGACATCACTCTTACCCAAATCAACAAGTCTTAAACATTACA 60

DB 457 CAATTTATCAAGATATATATCTTTGGCTAAATATCAATGAAACCTTAAACCCACC 516

QY 61 AGAAGCTATCCCTGAGAGAGAGAGAGAGAGTGTTCACACGAGAGATCTAGTGTGG 576

DB 517 TGAAGATGTCAACAGAGAGAGAGAGAGAGTGTTCACACGAGAGATCTAGTGTGG 636

QY 121 CAAGTCCCTCCCTCTAATTCCTCCCTAGTATCAATCTGGAGAGAGAGAGAGAGAG 180

DB 577 CAAATCTCTCCCTCTACCTCCCTATGATTTCTTGTGGAGAGAGAGAGAGAGAG 636

QY 181 CATTATCTACCCACTGC 201

DB 637 AATCTCTCTACCCAGACTGC 657

RESULT 12

BG013345

LOCUS BG013345 138 bp mRNA linear EST 24-JAN-2001

DEFINITION CM4-GN0361-161200-628-c08 GN0361 Homo sapiens cDNA, mRNA sequence.

ACCESSION BG013345

VERSION BG013345.1 GI:12463442

KEYWORDS

EST.

SOURCE

human.

ORGANISM

Homo sapiens

REFERENCE	Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS	Eukaryota; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 138)
TITLE	Dias Neto, E., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Nagai, M.A., Carvalho, A.F., Matsukuma, A., Bata, G.S., Simpson, D.H., Goldman, G.H., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare Brumsted, A.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.
JOURNAL	Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
MEDLINE	Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000) 2020263
COMMENT	Contact: Simpson A.J.G. Laboratory of Cancer Genetics Ludwig Institute for Cancer Research Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil Tel: +55-11-2704922 Fax: +55-11-2707001 Email: asimpson@ludwig.org.br This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL <a href="http://www.ludwig.org.br/scripts/gethtml2.pl?tl=CM&amp;t2=CM-GN0361-161200-628-c08&amp;t3=2000-12-16&amp;t4=1">http://www.ludwig.org.br/scripts/gethtml2.pl?tl=CM&amp;t2=CM-GN0361-</a> <a href="http://www.ludwig.org.br/scripts/gethtml2.pl?tl=CM&amp;t2=CM-GN0361-161200-628-c08&amp;t3=2000-12-16&amp;t4=1">161200-628-c08&amp;t3=2000-12-16&amp;t4=1</a> ) Seq primer: puc 18 forward High quality sequence start: 11 High quality sequence stop: 138. Location/Qualifiers 1..138 /organism="Homo sapiens" /db_xref="taxon:9606" /cclone_lib="GN0361" /dev_stage="Adult" /note="Organ: placenta.normal; Vector: puc18; Site:1: SmalI site:2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
FEATURES	
SOURCE	
BASE COUNT	42 a               37 c               22 g               37 t
ORIGIN	
Query Match	43.0%; Score 86.4; DB 12; Length 138;
Best Local Similarity	98.0%; Pred. No. 3e-16;
Matches	98; Conservative 0; Mismatches 1; Indels 1; Gaps 1;
OY	49 TAAACATTACAGACCTTCCCTGGAAGAAGGAAAGACTATTCACCCTTGCA 108       Db 27 TTAAACATTACAGGAA-CTATCCTCGAAGAGAGGAAAGACTATTCACCCTTGCA 85       OY 109 CATGGTATTAGTCAAGTCCCTCCCTCAATTCGCCCATCC 148       Db 86 CATGGTATTAGTCAAGTCCCTCCCTCAATTCGCCCATCC 125 
RESULT 13	
B54046	578 bp DNA linear GSS 20-JUN-1998
LOCUS	CIT-HSP-2014H6.TF CIT-HSP Homo sapiens genomic clone 2014H6, DNA sequence.
ACCESSION	B54046
VERSION	B54046.1 GI:2608380
KEYWORDS	GSS.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 578)
AUTHORS	Adams, M.D., Rounsley, S.D., Field, C.E., Bass, S., Linher, K., Golden 'K.', Berry, K., Granger, D., Sun, E., Wible, C., Shizuya, H., Simon, M. and Venter, J.C.

JOURNAL	Use of a random BAC End Sequence Database for Sequence-Ready Map Building									
COMMENT	Unpublished (1997) Other-GSSs: CIT-HSP-2014H6.TR Contact: Mark Adams Department of Eukaryotic Genomics The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850, USA Tel.: 301 838 0200 Fax: 301 838 0208 Email: mdadams@ligr.org Clones are available from Research Genetics (info@resgen.com): BAC end search page: http://www.ligr.org/tldb/hungen/bac_end_search/bac_end_search.html Seq primer: M13-21 Class: BAC ends.									
FEATURES	Location/Qualifiers									
source	1..578									
	/organism="Homo sapiens"									
	/db_xref="GDB:7042736"									
	/db_xref="taxon:9606"									
	/clone="2014H6"									
	/clone_lib="CIT-HSP"									
	/sex="Male"									
	/cell_type="Sperm"									
	/note="Vector: pheloBAC11; site_1: HindIII; site_2: HindIII"									
BASE COUNT	179 a 141 c 112 g 146 t									
ORIGIN										
Query Match	42.1%; Score 84.6; DB 17; Length 578;									
Best Local Similarity	67.2%; Pred. No. 1.8e-15;									
Matches 135; Conservative	0; Mismatches 64; Indels 2; Gaps 1;									
QY	1 CAACCTAGTTCAGACATCACCTCCTCTTGCCAAATACAAAGTTCTTAACATTACA 60									
Db	317 CAACCTGGTCAAAAGTATATACCTCTTGTGGCAAAATATACAGAACTTTAAAAACCTAC 376									
QY	61 AGGAACCTATCCTCCTAGAGAGAGGAAAGAACTATTCACCTTGTGACATGGTATTAGT 120									
Db	377 TGAATGATGTCAACAGAGAAAGGAGAGAGTGTTCACACAGAGATCTAGTGTGT 436									
QY	121 CAAGTCCTTCCTCTTAATTCCTCCATCCCTAGATACATCTGGGAGAGACCTTACCAGT 180									
Db	437 CAAATCTCTCTCTCTACCTCCCAT--CTATGAGCTCTGTGGGAAGAACCATCTACTAGT 494									
QY	181 CATTTATCTACCCCACTGC 201									
Db	495 AATCCTCTCTACCCCACTGC 515									
RESULT 14										
LOCUS	AQ427133 563 bp DNA linear GSS 24-MAR-1999									
DEFINITION	CITBI-E1-2566E24.TF CITBI-E1 Homo sapiens genomic clone 2566E24, DNA sequence.									
ACCESSION	AQ427133									
VERSION	AQ427133.1 GI:4499769									
KEYWORDS	GSS.									
SOURCE	human.									
ORGANISM	Homo sapiens									
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.									
AUTHORS	Zhao,S., Adams,M.D., Niernan,W., Malek,J., Shizuya,H., Simon,M. and Venter,J.C.									
TITLE	Map Building									
JOURNAL	Unpublished (1997)									
COMMENT	Other-GSSs: CITBI-E1-2566E24.TR Contact: Shaying Zhao, William Niernan, Mark Adams Department of Eukaryotic Genomics The Institute for Genomic Research									

9712 Medical Center Dr., Rockville, MD 20850

Tel: 301 838 0200

Fax: 301 838 0208

Email: hbeetlgr.org

Clones are available from Research Genetics (info@resgen.com). BAC

end search page:

http://www.tlgr.org/tldb/humgen/bac\_end\_search/bac\_end\_search.html.

Seq primer: M13-21

Class: BAC ends.

Location/Qualifiers

1..563

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="2566E24"

/clone\_1lb="CITBI-E1"

/sex="male"

/cell\_type="sperm"

/note="Vector: pBelOBAC11; Site\_1: EcoRI; Site\_2: EcoRI;

Caltech Human BAC Library D"

BASE COUNT 148 a 168 c 95 g 153 t

ORIGIN

Query Match

Best Local Similarity 90.7%; Pred. No. 7,3e-15; Length 563;

Matches 88; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

OY 105 GTGACATGTTATTCAGACCTCCCTTCCTTAATTCCTCCATCCCTAGATACATCTGG 164

DB 2 GTACATGTTATTCAGACCTCCCTTCCTTAATTCCTCCATCCCTAGATACATCTGG 61

OY 165 AAGGACCTACCCAGTATTTATCTACCCCAACTGG 201

DB 62 AAGGACCTACCCAGTATTTATCTACCCCAACTGG 98

RESULT 15

AO309705

LOCUS

DEFINITION CIT-HSP-2384F10.TFB CIT-HSP Homo sapiens genomic clone 2384F10, DNA

sequence.

ACCESSION

AO309705

VERSION

AO309705.1

KEYWORDS

GSS.

SOURCE

ORGANISM

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1 (bases 1 to 477)

Adams, M.D., Rounsley, S.D., Zhao, S., Bass, S., Linher, K.,

Berry, K., Granger, D., Suh, E., Wild, J.C., Shizuya, H., Simon, M. and

Venter, J.C.

Use of a random human BAC End Sequence Database for Sequence-Ready

Map Building

Unpublished (1998)

Other GSS: CIT-HSP-2384F10.TFB

Contact: Shaying Zhao, William Nierman, Mark Adams

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850

Tel: 301 838 0200

Fax: 301 838 0208

Email: hbeetlgr.org

Clones are available from Research Genetics (info@resgen.com). BAC

end search page:

http://www.tlgr.org/tldb/humgen/bac\_end\_search/bac\_end\_search.html.

Seq primer: M13-21

Class: BAC ends.

Location/Qualifiers

1..477

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="2384F10"

/clone\_1lb="CIT-HSP"

/sex="Male"

/cell\_type="sperm"

/note="Vector: pBelOBAC11; Site\_1: HindIII; Site\_2:

HindIII"

BASE COUNT 153 a 115 c 93 g 116 t

ORIGIN

Query Match

Best Local Similarity 67.3%; Pred. No. 4.4e-14; Length 477;

Matches 113; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

OY 1 CAACCTAGTGCAGACATCACCTCCTTAGCCCAATATCAACAAGTTCTTAACATTTACA 60

DB 310 CAACCTAGTGCAGACATCACCTCCTTAGCCCAATATCAACAAGTTCTTAACATTTACA 369

OY 61 AGGACCTATCCCTGAG 120

DB 370 TGAAGCATGTCACAG 429

OY 121 CAAGTCCCTCCCTCTAATTCCTCCATCCCTAGATATCATCTGGAAGG 168

DB 430 CAATCTCTCCCTCTAATTCCTCCATCCCTAGATATCATCTGGAAGG 477

Search completed: May 2, 2003, 16:21:29

Job time : 1117 secs

GenCore version 5.1.5  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 2, 2003, 14:52:21 ; Search time 39.333 Seconds  
(without alignments)  
1567.170 Million cell updates/sec

Title: US-09-719-554-3\_COPY\_7500\_7700

Perfect score: 201  
Sequence: 1 caactagttgcagacatca.....atttaccacccactgc 201

Scoring table: IDENTITY\_NUC  
Gapop 10.0 ; Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

Issued\_Patents\_NA.\*  
1: /cgn2\_6/ptodata/2/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/2/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/2/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/2/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/2/ina/PCTUS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/2/ina/Backfile1.seq.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	201	100.0	2946	4	US-09-175-928-3
2	146.6	72.9	542	1	US-08-686-878A-48
3	146.6	72.9	542	1	US-08-721-489-2
4	29.2	14.5	50000	4	US-09-146-053-4
5	28.6	14.2	1564	4	US-09-071-035-63
6	28.6	14.2	1641	4	US-09-071-035-61
7	28.4	14.1	735	1	US-08-798-000-3
8	28.4	14.1	1234	1	US-08-798-000-1
9	28.4	14.1	2431	3	US-08-985-526-35
10	28.4	14.1	3612	4	US-09-265-013-3
11	28.4	14.1	5406	1	US-07-813-593-5
12	28.4	14.1	5406	1	US-07-977-451-5
13	28.4	14.1	5406	1	US-07-946-507-3
14	28.4	14.1	5406	1	US-08-252-517-5
15	28.4	14.1	5406	1	US-07-906-397A-5
16	28.4	14.1	5406	1	US-08-601-891-5
17	28.4	14.1	5406	2	US-09-021-324-5
18	28.4	14.1	5406	5	PCT-US92-02750-7
19	28.4	14.1	5406	5	PCT-US92-05401-5
20	28.4	14.1	5406	5	PCT-US92-09893-5
21	28.4	14.1	5470	2	US-08-443-861-1
22	28.4	14.1	5470	4	US-08-193-829B-1
23	28.4	14.1	6028	4	US-09-011-745-5
24	28.4	14.1	6061	4	US-09-011-745-6
25	28.4	14.1	6350	2	US-08-385-335A-8
26	28.4	14.1	7308	4	US-09-011-745-3
27	28.4	14.1	7308	4	US-09-011-745-4

28	28.4	14.1	7616	4	US-09-011-745-2	Sequence 2, Appli
29	28.4	14.1	8332	3	US-08-850-961-1	Sequence 1, Appli
30	28.4	14.1	8332	4	US-09-479-776-1	Sequence 1, Appli
31	28.4	14.1	8332	4	US-09-309-572-11	Sequence 11, Appli
32	28.4	14.1	8332	4	US-09-315-127-1	Sequence 1, Appli
33	28.4	14.1	8332	4	US-09-265-013-1	Sequence 1, Appli
34	28.4	14.1	8614	4	US-09-133-827-5	Sequence 5, Appli
35	28.4	14.1	10100	4	US-09-133-944-1	Sequence 1, Appli
36	28.4	14.1	10100	4	US-09-208-827-1	Sequence 1, Appli
37	28	13.9	568	4	US-09-404-879A-113	Sequence 113, App
38	27.8	13.8	11703	3	US-09-101-886B-3	Sequence 3, Appli
39	27.4	13.6	5057	2	US-08-365-486A-12	Sequence 12, Appli
40	27.4	13.6	5057	4	US-08-880-342-12	Sequence 12, Appli
41	27.4	13.6	5108	1	US-07-642-002-1	Sequence 1, Appli
42	27.4	13.6	10367	1	US-08-110-300A-9	Sequence 9, Appli
43	27.4	13.6	10367	2	US-08-886-642-9	Sequence 9, Appli
44	27.4	13.6	10367	5	PCT-US93-08041-9	Sequence 9, Appli
45	27.2	13.5	771	4	US-08-998-416-222	Sequence 222, App

## ALIGNMENTS

```
RESULT 1
US-09-175-928-3
; Sequence 3, Application US/09175928A
; Patent No. 6312921
;
GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John M.
; APPLICANT: Lavallie, Edward R.
; APPLICANT: Collins-Racie, Lisa A.
; APPLICANT: Evans, Cheryl
; APPLICANT: Merberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Mi, Sha
; APPLICANT: Genetics Institute, Inc.
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM
; FILE REFERENCE: 6006B.A172A
; CURRENT APPLICATION NUMBER: US/09/175,928A
; CURRENT FILING DATE: 1998-10-20
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 3
; LENGTH: 2946
; TYPE: DNA
; ORGANISM: Homo sapiens
;
Query Match          100.0%; Score 201; DB 4; Length 2946;
Best Local Similarity 100.0%; Pred. No. 6.2e-55;
Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
;
QY 1 CAACTAGTTGGAGACATCCTCCTAGCCAAATATCAACAAGTCTTAAACATTACA 60
DB 549 CAACTAGTTGGAGACATCCTCCTAGCCAAATATCAACAAGTCTTAAACATTACA 608
;
QY 61 AGGAACCTATCCTCTGAGAGAGAGAAAGAACTATTCACCTTGTGAATGTTATTACT 120
DB 609 AGGAACCTATCCTCTGAGAGAGAGAAAGAAAGAACTATTCACCTTGTGAATGTTATTACT 668
;
QY 121 CAAGTCCCTCCCTCAATTCCTCCATCCCTAGATACATCTGGGAAGACCTACCCACT 180
DB 669 CAAGTCCCTCCCTCAATTCCTCCATCCCTAGATACATCTGGGAAGACCTACCCACT 728
;
QY 181 CATTATCTACCCCACTGC 201
DB 729 CATTATCTACCCCACTGC 749
;
RESULT 2
US-08-686-878A-48
; Sequence 48, Application US/08686878A
```

Patent No. 5708157  
GENERAL INFORMATION:  
APPLICANT: Jacobs, Kenneth  
APPLICANT: McCoy, John  
APPLICANT: Lavallee, Edward  
APPLICANT: Racie, Lisa  
APPLICANT: Merberg, David  
APPLICANT: Treacy, Maurice  
APPLICANT: Evans, Cheryl  
APPLICANT: Spaulding, Vikki  
TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES  
NUMBER OF SEQUENCES: 71  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genetics Institute, Inc.  
STREET: 87 Cambridgepark Drive  
CITY: Cambridge  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02140  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/686,878A  
FILING DATE:  
CLASSIFICATION: 536  
ATTORNEY/AGENT INFORMATION:  
NAME: Brown, Scott A.  
REGISTRATION NUMBER: 32,724  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 498-8224  
TELEFAX: (617) 876-5851  
INFORMATION FOR SEQ. ID NO.: 48:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 542 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
US-08-686-878A-48  
Query Match 72.9%; Score 146.6; DB 1; Length 542;  
Best Local Similarity 91.4%; Pred. No. 8, 1e-38;  
Matches 149; Conservative 0; Mismatches 14; Indels 0; Gaps 0;  
QY 39 AACAGTCTTTAAACATTACAGAACCTATCCCTGAGAGAGGAGGAGAAAGAACTATTTC 98  
DB 1 AACAAATTTTAAACATTACAGAACGATCCNTGAGNAGGAGGAGAAAGAAATTTATTC 60  
QY 99 ACCCTTGACATGATATTAGTCAAGTCCCTCTAATTCCTCCATCCCTAGATACAT 158  
DB 61 ACCCTTGACATGATATTAGTCAAGTCCCTCTAATTCCTCCATCCCTAGATACAT 120  
QY 159 CCGGGAAGAGACCTACCCAGTATTTTATCTACCCCACTGC 201  
DB 121 CNTGGAGGAGCCCTACCCAGTATTTTATNTATACCCCACTGC 163  
RESULT 3  
US-08-721-489-2  
Sequence 2, Application US/08721489  
Patent No. 5786465  
GENERAL INFORMATION:  
APPLICANT: Jacobs, Kenneth  
APPLICANT: McCoy, John  
APPLICANT: Lavallee, Edward  
APPLICANT: Racie, Lisa  
APPLICANT: Merberg, David  
APPLICANT: Treacy, Maurice  
APPLICANT: Spaulding, Vikki

TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genetics Institute, Inc.  
STREET: 87 Cambridgepark Drive  
CITY: Cambridge  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02140  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/721,489  
FILING DATE:  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Brown, Scott A.  
REGISTRATION NUMBER: 32,724  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 498-8224  
TELEFAX: (617) 876-5851  
INFORMATION FOR SEQ. ID NO.: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 542 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
US-08-721-489-2  
Query Match 72.9%; Score 146.6; DB 1; Length 542;  
Best Local Similarity 91.4%; Pred. No. 8, 1e-38;  
Matches 149; Conservative 0; Mismatches 14; Indels 0; Gaps 0;  
QY 39 AACAGTCTTTAAACATTACAGAACCTATCCCTGAGAGAGGAGGAGAAAGAACTATTTC 98  
DB 1 AACAAATTTTAAACATTACAGAACGATCCNTGAGNAGGAGGAGAAAGAAATTTATTC 60  
QY 99 ACCCTTGACATGATATTAGTCAAGTCCCTCTAATTCCTCCATCCCTAGATACAT 158  
DB 61 ACCCTTGACATGATATTAGTCAAGTCCCTCTAATTCCTCCATCCCTAGATACAT 120  
QY 159 CCGGGAAGAGACCTACCCAGTATTTTATCTACCCCACTGC 201  
DB 121 CNTGGAGGAGCCCTACCCAGTATTTTATNTATACCCCACTGC 163  
RESULT 4  
US-09-146-053-4  
Sequence 4, Application US/09146053A  
Patent No. 6399349  
GENERAL INFORMATION:  
APPLICANT: Ryan, James W.  
APPLICANT: Spinkie, Jerry Joe Curtis  
APPLICANT: Venema, Richard C.  
TITLE OF INVENTION: Human Aminopeptidase P Gene  
FILE REFERENCE: MCG103  
CURRENT APPLICATION NUMBER: US/09/146,053A  
EARLIER FILING DATE: 1998-09-02  
EARLIER APPLICATION NUMBER: 60/057,854  
EARLIER FILING DATE: 1997-09-02  
NUMBER OF SEQ. ID NOS: 7  
SOFTWARE: PatentIn Ver. 2.0  
SEQ. ID NO. 4  
LENGTH: 50000  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-146-053-4



CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/011,564  
FILING DATE: 13-FEB-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Parmelee, Steven W.  
REGISTRATION NUMBER: 31,990  
REFERENCE/DOCKET NUMBER: 14538A-002010  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 206-467-9600  
TELEFAX: 415-576-0300  
INFORMATION FOR SEQ. ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 735 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: both  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
US-08-798-000-3

Query Match 14.1% Score 28.4; DB 1; Length 735;  
Best Local Similarity 70.4%; Pred. No. 4.3;  
Matches 38; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 148 CCTAGATACATCTGGGAAGACCTTACCCAGTATTATCTACCCCACTGC 201  
DB 500 CCTAGAACCTCGCTGGAAAGACCTTACACAGTCTGTGACACCCCAACCCG 553

## RESULT 8

US-08-798-000-1  
Sequence 1, Application US/08798000  
Patent No. 5766945

GENERAL INFORMATION:  
APPLICANT: Miller, A. Dusty  
TITLE OF INVENTION: 10A1 RETROVIRAL PACKAGING CELL LINES  
NUMBER OF SEQUENCES: 3  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Center, 8th floor  
CITY: San Francisco  
STATE: CA  
COUNTRY: USA  
ZIP: 94111-3834  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/798,000  
FILING DATE: 12-FEB-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/011,564  
FILING DATE: 13-FEB-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Parmelee, Steven W.  
REGISTRATION NUMBER: 31,990  
REFERENCE/DOCKET NUMBER: 14538A-002010  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 206-467-9600  
TELEFAX: 415-576-0300  
INFORMATION FOR SEQ. ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1234 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: both  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
US-08-798-000-1

Query Match 14.1% Score 28.4; DB 1; Length 1234;

Best Local Similarity 70.4%; Pred. No. 5;  
Matches 38; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

QY 148 CCTAGATACATCTGGGAAGACCTTACCCAGTATTATCTACCCCACTGC 201  
DB 500 CCTAGAACCTCGCTGGAAAGACCTTACACAGTCTGTGACACCCCAACCCG 553

## RESULT 9

US-08-985-526-35  
Sequence 35, Application US/08985526  
Patent No. 6080728

GENERAL INFORMATION:  
APPLICANT: Mixson, James A.  
TITLE OF INVENTION: CARRIER-DNA COMPLEXES CONTAINING DNA  
TITLE OF INVENTION: ENCODING ANTI-ANGIOGENIC PEPTIDES AND THEIR USE IN GENE  
TITLE OF INVENTION: THERAPY  
NUMBER OF SEQUENCES: 43  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Connolly, Bove, Lodge, & Hutz  
STREET: 1220 Market Street, P.O. Box 2207  
CITY: Wilmington  
STATE: Delaware  
COUNTRY: U.S.A.  
ZIP: 19899  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/985,526  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/608,845  
FILING DATE: 16-JUL-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: McMorris Jr., Robert G.  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (302) 658-9141  
TELEFAX: (302) 658-5613  
INFORMATION FOR SEQ. ID NO: 35:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2431 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-985-526-35

Query Match 14.1% Score 28.4; DB 3; Length 2431;  
Best Local Similarity 54.9%; Pred. No. 6.1;  
Matches 56; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

QY 9 TTGCACATACATCTGCTAGCCCAATATACCAAGTCTTAAACATTACAGAACCT 68  
DB 791 TGCACTCTCCACCTCAAGTCTCATATAGAAAGATTGTAAACCGGATGGAACCC 850

QY 69 ATCCCTGAGAAGAGGAAAGAACTATTCACACCTTGACAA 110  
DB 851 TTTCCTGGACGTGGCGAAGATGTTTGGACGACCTTGACA 892

## RESULT 10

US-09-265-013-3  
Sequence 3, Application US/09265013  
Patent No. 6451304

GENERAL INFORMATION:  
APPLICANT: FRIEDMAN, Theodore  
TITLE OF INVENTION: METHOD FOR RETROVIRUS VECTOR PRODUCTION BY SEPARATED  
TITLE OF INVENTION: GAG AND POL EXPRESSION  
FILE REFERENCE: 041673/2010



CURRENT APPLICATION NUMBER: US/09/265.013  
CURRENT FILING DATE: 1999-03-09  
NUMBER OF SEQ ID NOS: 5  
SOFTWARE: Patent In Ver. 2.1  
SEQ ID NO 3  
LENGTH: 3612  
TYPE: DNA  
ORGANISM: Moloney murine leukemia virus (MOLV)  
US-09-265-013-3

Query Match 14.1%, Score 28.4; DB 4; Length 3612;  
Best Local Similarity 70.4%; Pred. No. 6.9;  
Matches 38; Conservative 0; Mismatches 16; Indels 0; Gaps 0;

OY 148 CCTGATCATCTCGGAGAGCCTACCGCTCTTTATCTACCCACCTGC 201  
DB 3414 CCTAGACCTCGCTGGAAGACCTTACACAGCTCTGCTGACCCACCGC 3467

RESULT 11  
US-07-813-593-3  
Sequence 3, Application US/07813593  
Patent No. 5183438  
GENERAL INFORMATION:  
APPLICANT: Lemischka, Ihor R.  
TITLE OF INVENTION: TOTIPOTENT HEMATOPOIETIC STEM CELL  
TITLE OF INVENTION: RECEPTORS AND THEIR LIGANDS  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: IMCONE SYSTEMS INCORPORATED  
STREET: 180 VARICK STREET  
CITY: NEW YORK  
STATE: NEW YORK  
COUNTRY: U.S.A.  
ZIP: 10014  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/813.593  
FILING DATE: 19920415  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/793.065  
FILING DATE: 15-NOV-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/728.913  
FILING DATE: 28-JUN-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/679.666  
FILING DATE: 02-APR-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Felt, Irving N.  
REGISTRATION NUMBER: 28.601  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-645-1405  
TELEFAX: 212-645-2054  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 5406 base pairs  
TYPE: NUCLEIC ACID  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 208..4311  
FEATURE:  
NAME/KEY: mat\_peptide  
LOCATION: 208..4308

US-07-813-593-3

Query Match 14.1%, Score 28.4; DB 1; Length 5406;  
Best Local Similarity 54.9%; Pred. No. 7.8;  
Matches 56; Conservative 0; Mismatches 46; Indels 0; Gaps 0;

OY 9 TTGCAGACATCACCTCTCTAGCCAAATATACAGATTTCTTAACATTACAGAACCT 68  
DB 991 TTGCACCTCTCCACCTTCAAGATCTCATATAGAAAGATTGTAAACCGGATGTGAACCC 1050

OY 69 ATCCCTGAGAGAGGAGAAAGACTATTCACCCCTTGACAA 110  
DB 1051 TTTCCTGGGACTGTGGCAGATGTTTTGACCACTTGACA 1092

RESULT 12  
US-07-977-451-5  
Sequence 5, Application US/07977451  
Patent No. 5270458  
GENERAL INFORMATION:  
APPLICANT: Lemischka, Ihor R.  
TITLE OF INVENTION: TOTIPOTENT HEMATOPOIETIC STEM CELL  
TITLE OF INVENTION: RECEPTORS AND THEIR LIGANDS  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Imclone Systems Incorporated  
STREET: 180 Varick Street  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10014  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/977.451  
FILING DATE: 19921119  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US UNASSIGNED  
FILING DATE: 12-NOV-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/906.397  
FILING DATE: 26-JUN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US PCT/US92/05401  
FILING DATE: 26-JUN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: TW 81102961  
FILING DATE: 15-APR-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US PCT/US92/02750  
FILING DATE: 02-APR-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/813.593  
FILING DATE: 24-DEC-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/793.065  
FILING DATE: 15-NOV-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/728.913  
FILING DATE: 28-JUN-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/679.666  
FILING DATE: 02-APR-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Felt, Irving N.  
REGISTRATION NUMBER: 28.601  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-645-1405

```

1 TELEFAX: 212-645-2054
2 INFORMATION FOR SEQ ID NO: 5
3 SEQUENCE CHARACTERISTICS:
4 LENGTH: 5406 base pairs
5 TYPE: NUCLEIC ACID
6 STRANDEDNESS: double
7 TOPOLOGY: linear
8 MOLECULE TYPE: cDNA
9 HYPOTHEICAL: NO
10 ANTI-SENSE: NO
11 FRAGMENT TYPE: N-terminal
12 FEATURE:
13 NAME/KEY: CDS
14 LOCATION: 208..4311
15 FEATURE:
16 NAME/KEY: mat_peptide
17 LOCATION: 265..4308
18 FEATURE:
19 NAME/KEY: sig_peptide
20 LOCATION: 208..264
21
22 NS-07-977-451-5

```

Query Match	14.1%	Score 28.4	DB 1	Length 5406
Best Local Similarity	54.9%	Pred. No. 7.8		
Matches 56; Conservative	0	Mismatches 46	Indels 0	Gaps 0

Oy 9 TTCCAGACATCACCCTCTTAGCCAAATATCAAGAAGTCTTAACAATTCAGAAGAACCT 68  
| | | | | | | | | | | | | | | | | |  
Db 991 TGGCACATCCACCTTCAAGTCTCATATGAAGAAGTTGTAAACCGGATGTCAAACC 1050

QY 69 ATCCCTGAGAGAGGGAAAGAATAATTCACCCCTTGTGACA 110  
| | | | |  
Db 1051 TTCTCGGCACGTGGGCCAAGATGTTTTTGACGACCCTTGACA 109

RESULT 13  
 US-07-946-507-3  
 Sequence 3, Application US/07946507  
 Patent No. 5283354  
 GENERAL INFORMATION:  
 APPLICANT: Lemischka, Thor R.  
 TITLE OF INVENTION: TOTIPOTENT HEMATOPOIETIC STEM CELL  
 TITLE OF INVENTION: TOTIPOTENT HEMATOPOIETIC STEM CELL  
 NUMBER OF INVENTION: RECEPTORS AND THEIR LIGANDS  
 NUMBER OF SEQUENCES: 4  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: IMCLONE SYSTEMS INCORPORATED  
 STREET: 180 VARICK STREET  
 CITY: NEW YORK  
 STATE: NEW YORK  
 COUNTRY: U.S.A.  
 ZIP: 10014  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/07/946,507  
 FILING DATE: 19920917  
 CLASSIFICATION: 536  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US/07/813,593  
 FILING DATE: 24-DEC-1991  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US/07/793,065  
 FILING DATE: 15-NOV-1991  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US/07/728,913  
 FILING DATE: 28-JUN-1991  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US/07/679,666  
 FILING DATE: 02-APR-1991  
 ATTORNEY/AGENT INFORMATION:

```

1 NAME: Felt, Irving N.
2 REGISTRATION NUMBER: 28,601
3 REFERENCE/DOCKET NUMBER: LEM-3-PPP
4 TELECOMMUNICATION INFORMATION:
5 TELEPHONE: 212-645-1405
6 TELEFAX: 212-645-2054
7 INFORMATION FOR SEQ ID NO: 3:
8 SEQUENCE CHARACTERISTICS:
9 LENGTH: 5406 base pairs
10 TYPE: NUCLEIC ACID
11 STRANDEDNESS: single
12 TOPOLOGY: linear
13 MOLECULE TYPE: cDNA
14 FEATURE:
15 NAME/KEY: CDS
16 LOCATION: 208..4311
17 FEATURE:
18 NAME/KEY: mat_peptide
19 LOCATION: 208..4508
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
10
```

Query Match	41.1%	Score	28.4	DB	1	Length	5406
Best Local	54.9%	Pred.	No. 7.8				
Matches	56	Conservative	0	Mismatches	46	Indels	0
						Gaps	0

QY 9 TTGCAGACATCACCTCCTTAGCCAAATATCAACAAGTCTTAACAATTACAGAACCCT 68  
- - - - -  
Db 991 TGGCACTCTCCACCTTCAAAGTCTCATATGAGAAGTTGTAAACCGGATGTGAALACC 1050

QY 69 ATCCCTGAGAGAGGGAAAAGAATACTATTCACCCCTGTGACA 110  
| | | | | | | | | | | | | |  
Db 1051 TTTCCTGGGACTGTGGCGAAGAGTATTGTGACCACTTGACA 109

RESULT 14  
 US-08-252-517-5  
 Sequence 5, Application US/08252517  
 Patent No. 5348065  
 GENERAL INFORMATION:  
 APPLICANT: Lemischka, Ihor R.  
 TITLE OF INVENTION: TOTIPOTENT HEMATOPOIETIC STEM CELLS  
 TITLE OF INVENTION: RECEPTORS AND THEIR LIGANDS  
 NUMBER OF SEQUENCES: 10  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Imclone Systems Incorporated  
 STREET: 180 Varick Street  
 CITY: New York  
 STATE: New York  
 COUNTRY: U.S.A.  
 ZIP: 10014  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/252,517  
 FILING DATE: 31-OCT-1994  
 CLASSIFICATION: 530  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/977,451  
 FILING DATE: 19-NOV-1992  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/906,397  
 FILING DATE: 26-JUN-1992  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US PCT/US92/05401  
 FILING DATE: 26-JUN-1992  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: TW 81102961  
 FILING DATE: 15-APR-1992  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US PCT/US92/02750





Gencore version 5.1.5  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 2, 2003, 16:21:41 ; Search time 63.333 Seconds

(without alignments)  
3750.092 Million cell updates/sec

Title: US-09-719-554-3\_COPY\_7500\_7700

Perfect score: 201  
Sequence: 1 caactagtgtgcagacatca.....atttattcaccacactgc 201

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 746064 seqs, 590810554 residues

Total number of hits satisfying chosen parameters: 1492128

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications -NA:\*  
1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:\*  
5: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:\*  
6: /cgn2\_6/ptodata/1/pubpna/PCRTUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*  
10: /cgn2\_6/ptodata/1/pubpna/US05\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:\*  
12: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq:\*  
13: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:\*  
14: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	201	100.0	2930	10	US-09-902-535-1 Sequence 1, Appl
2	201	100.0	2946	9	US-10-114-893-134 Sequence 134, Appl
3	146.6	72.9	542	12	US-10-040-916-48 Sequence 48, Appl
4	100.8	50.1	758	7	US-08-979-847-112 Sequence 112, Appl
5	46.4	23.1	200	10	US-09-864-761-20979 Sequence 20979, A
6	33.4	16.6	2482	10	US-09-764-853-198 Sequence 198, Appl
7	32.8	16.3	9990	10	US-09-767-088A-2 Sequence 2, Appl
8	32.8	16.3	9990	10	US-09-767-088A-15 Sequence 15, Appl
9	32.6	16.2	10098	9	US-09-764-868-1482 Sequence 1482, Ap
10	31.4	15.6	550	9	US-09-918-936-1452 Sequence 1452, Ap
11	30.8	15.3	509	9	US-09-918-936-11931 Sequence 11931, A
12	30.8	15.3	2456	9	US-09-822-846-42 Sequence 42, Appl
13	30.8	15.3	26197	9	US-10-114-893-122 Sequence 122, Appl
14	30.8	15.3	26197	9	US-10-092-154-1965 Sequence 1965, Ap
15	30.8	15.3	26197	9	US-10-092-154-1965 Sequence 1965, Ap
16	30.8	15.3	26210	9	US-10-092-154-1966 Sequence 1966, Ap
17	30.8	15.3	26210	10	US-09-764-847-1966 Sequence 1966, Ap
18	30.2	15.0	1691139	9	US-10-067-514-1 Sequence 1, Appl
19	30	14.9	1176	10	US-09-764-898-23 Sequence 23, Appl

20	29.6	14.7	159	9	US-09-728-444-16 Sequence 16, Appl
21	29.6	14.7	159	9	US-09-728-444-17 Sequence 17, Appl
22	29.6	14.7	159	9	US-09-728-444-18 Sequence 18, Appl
23	29.6	14.7	159	9	US-09-728-444-96 Sequence 96, Appl
24	29.6	14.7	159	9	US-09-728-444-98 Sequence 98, Appl
25	29.6	14.7	159	9	US-09-728-444-100 Sequence 100, Appl
26	29.6	14.7	258	9	US-09-535-459-1182 Sequence 1182, Appl
27	29.6	14.7	1913	9	US-10-102-806-1 Sequence 1, Appl
28	29.6	14.7	32192	9	US-09-764-891-7049 Sequence 7049, Ap
29	29.4	14.6	432	10	US-09-764-877-72511 Sequence 2511, Ap
30	29.2	14.5	66109	10	US-09-880-107-3768 Sequence 3768, Ap
31	29.2	14.5	106323	10	US-09-803-661-3 Sequence 3, Appl
32	29.2	14.5	198285	10	US-09-880-107-3814 Sequence 3814, Ap
33	29	14.4	339	9	US-09-803-719-2220 Sequence 2220, Ap
34	29	14.4	446	9	US-10-046-935-1774 Sequence 1774, Ap
35	29	14.4	446	9	US-09-878-178-1774 Sequence 1774, Ap
36	29	14.4	446	9	US-10-146-502-1774 Sequence 1774, Ap
37	29	14.4	501	10	US-09-920-300A-34 Sequence 34, Appl
38	29	14.4	501	12	US-10-033-528-34 Sequence 34, Appl
39	29	14.4	2292	10	US-09-898-598-2602 Sequence 2602, Ap
40	29	14.4	3472	10	US-09-873-737A-5 Sequence 5, Appl
41	28.8	14.3	208	9	US-09-796-692-7291 Sequence 7291, Ap
42	28.8	14.3	208	9	US-10-040-862-7291 Sequence 7291, Ap
43	28.8	14.3	493	10	US-09-783-590-8089 Sequence 8089, Ap
44	28.8	14.3	618	10	US-09-764-898-96 Sequence 96, Appl
45	28.8	14.3	775	9	US-09-796-692-7274 Sequence 7274, Ap

## ALIGNMENTS

RESULT 1	US-09-902-535-1	Sequence 1, Application US/09902535
1	US-09-902-535-1	Patent No. US20020102530A1
2	US-09-902-535-1	GENERAL INFORMATION:
3	US-09-902-535-1	APPLICANT: Kelch, Jr., James C.
4	US-09-902-535-1	APPLICANT: McCoy, John M.
5	US-09-902-535-1	APPLICANT: M. Sha
6	US-09-902-535-1	TITLE OF INVENTION: Methods and compositions for diagnosing
7	US-09-902-535-1	TITLE OF INVENTION: and treating preclampsia and gestational trophoblast
8	US-09-902-535-1	FILE REFERENCE: GIN-6006B4
9	US-09-902-535-1	CURRENT APPLICATION NUMBER: US/09/902,535
10	US-09-902-535-1	CURRENT FILING DATE: 2001-07-09
11	US-09-902-535-1	PRIOR APPLICATION NUMBER: 60/216,657
12	US-09-902-535-1	PRIOR FILING DATE: 2000-07-06
13	US-09-902-535-1	NUMBER OF SEQ ID NOS: 5
14	US-09-902-535-1	SOFTWARE: FastSeq for Windows Version 4.0
15	US-09-902-535-1	SEQ ID NO 1
16	US-09-902-535-1	LENGTH: 2930
17	US-09-902-535-1	TYPE: DNA
18	US-09-902-535-1	ORGANISM: Homo sapiens
19	US-09-902-535-1	FEATURE:
20	US-09-902-535-1	NAME/KEY: CDS
21	US-09-902-535-1	LOCATION: (930)...
22	US-09-902-535-1	Matches 201; Conservative 0; Indels 0; Gaps 0;
23	US-09-902-535-1	Best Local Similarity 100.0%; Pmed. No. 5.3e-56;
24	US-09-902-535-1	Query Match 100.0%; Score 201; DB 10; Length 2930;
25	US-09-902-535-1	1 CAACTAGTTCGACACATCACCCTCTTGCCCAATATATACAAAGTCTTAATAACATTACA 60
26	US-09-902-535-1	1 CACTTACTTCGACACATCACCCTCTTGCCCAATATATACAAAGTCTTAATAACATTACA 60
27	US-09-902-535-1	1 CACTTACTTCGACACATCACCCTCTTGCCCAATATATACAAAGTCTTAATAACATTACA 60
28	US-09-902-535-1	1 CACTTACTTCGACACATCACCCTCTTGCCCAATATATACAAAGTCTTAATAACATTACA 60
29	US-09-902-535-1	1 CACTTACTTCGACACATCACCCTCTTGCCCAATATATACAAAGTCTTAATAACATTACA 60
30	US-09-902-535-1	1 CACTTACTTCGACACATCACCCTCTTGCCCAATATATACAAAGTCTTAATAACATTACA 60
31	US-09-902-535-1	1 CACTTACTTCGACACATCACCCTCTTGCCCAATATATACAAAGTCTTAATAACATTACA 60
32	US-09-902-535-1	1 CACTTACTTCGACACATCACCCTCTTGCCCAATATATACAAAGTCTTAATAACATTACA 60
33	US-09-902-535-1	1 CACTTACTTCGACACATCACCCTCTTGCCCAATATATACAAAGTCTTAATAACATTACA 60
34	US-09-902-535-1	1 CACTTACTTCGACACATCACCCTCTTGCCCAATATATACAAAGTCTTAATAACATTACA 60
35	US-09-902-535-1	1 CACTTACTTCGACACATCACCCTCTTGCCCAATATATACAAAGTCTTAATAACATTACA 60
36	US-09-902-535-1	1 CACTTACTTCGACACATCACCCTCTTGCCCAATATATACAAAGTCTTAATAACATTACA 60
37	US-09-902-535-1	1 CACTTACTTCGACACATCACCCTCTTGCCCAATATATACAAAGTCTTAATAACATTACA 60
38	US-09-902-535-1	1 CACTTACTTCGACACATCACCCTCTTGCCCAATATATACAAAGTCTTAATAACATTACA 60
39	US-09-902-535-1	1 CACTTACTTCGACACATCACCCTCTTGCCCAATATATACAAAGTCTTAATAACATTACA 60
40	US-09-902-535-1	1 CACTTACTTCGACACATCACCCTCTTGCCCAATATATACAAAGTCTTAATAACATTACA 60
41	US-09-902-535-1	1 CACTTACTTCGACACATCACCCTCTTGCCCAATATATACAAAGTCTTAATAACATTACA 60
42	US-09-902-535-1	1 CACTTACTTCGACACATCACCCTCTTGCCCAATATATACAAAGTCTTAATAACATTACA 60
43	US-09-902-535-1	1 CACTTACTTCGACACATCACCCTCTTGCCCAATATATACAAAGTCTTAATAACATTACA 60
44	US-09-902-535-1	1 CACTTACTTCGACACATCACCCTCTTGCCCAATATATACAAAGTCTTAATAACATTACA 60
45	US-09-902-535-1	1 CACTTACTTCGACACATCACCCTCTTGCCCAATATATACAAAGTCTTAATAACATTACA 60

OY 181 CATTATCTACCCCACTGC 201  
|||||  
DB 731 CATTATCTACCCCACTGC 751

## RESULT 2

US-10-114-893-134  
Sequence 134, Application US/10114893  
Publication No. US20020193567A1  
GENERAL INFORMATION:

APPLICANT: Jacobs, Kenneth  
APPLICANT: MCCOY, John M.  
APPLICANT: LAVAILLE, Edward R.  
APPLICANT: COLLINS-RACIE, Lisa A.  
APPLICANT: Evans, Cheryl  
APPLICANT: Merberg, David  
APPLICANT: Treacy, Maurice  
APPLICANT: Bowman, Michael R.  
APPLICANT: Spaulding, Vikki  
APPLICANT: Carlin-Duckett, McKeough  
APPLICANT: Kelleher, Kerry S.  
APPLICANT: Genetics Institute, Inc.  
TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM  
FILE REFERENCE: GI 6000-10A  
CURRENT APPLICATION NUMBER: US/10/114,893  
CURRENT FILING DATE: 2002-04-02  
EARLIER APPLICATION NUMBER: 09/413,232  
EARLIER FILING DATE: 1999-10-06  
NUMBER OF SEQ ID NOS: 321  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 134  
LENGTH: 2946  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-10-114-893-134

## Query Match

Best Local Similarity 100.0%; Score 201; DB 9; Length 2946;  
Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CAACCTAGTTCGACATCTCTCTAGCCCAATTCACACAGTCTTAAACATTACA 60  
|||||  
DB 549 CAACCTAGTTCGACATCTCTCTAGCCCAATTCACACAGTCTTAAACATTACA 608  
OY 61 AGGAACCTATCCCTGAGAGAGGAAAGAAAGAACTATTCACACCTTGTGACATGTTAGT 120  
|||||  
DB 609 AGGAACCTATCCCTGAGAGAGGAAAGAAAGAACTATTCACACCTTGTGACATGTTAGT 668  
OY 121 CAAGTCCCTTCCCTCTAATTCCTCCATCCCTATATCATCTGGAGAGGACCTACCACT 180  
|||||  
DB 669 CAAGTCCCTTCCCTCTAATTCCTCCATCCCTATATCATCTGGAGAGGACCTACCACT 728  
OY 181 CATTATCTACCCCACTGC 201  
|||||  
DB 729 CATTATCTACCCCACTGC 749

## RESULT 3

US-10-040-916-48  
Sequence 48, Application US/10040916  
Patent No. US20020146769A1  
GENERAL INFORMATION:

APPLICANT: Jacobs, Kenneth  
APPLICANT: MCCOY, John  
APPLICANT: LAVAILLE, Edward  
APPLICANT: RACIE, Lisa  
APPLICANT: Merberg, David  
APPLICANT: Treacy, Maurice  
APPLICANT: Evans, Cheryl  
APPLICANT: Spaulding, Vikki  
TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES  
ENCODING THEM

NUMBER OF SEQUENCES: 71  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genetics Institute, Inc.  
STREET: 87 Cambridgepark Drive  
CITY: Cambridge  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02140

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/10/040,916  
FILING DATE: 07-Jan-2002  
CLASSIFICATION: <unknown>

## PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/887,029  
FILING DATE: 07-FEB-1997  
APPLICATION NUMBER: 08/686,878  
FILING DATE: <unknown>

## ATTORNEY/AGENT INFORMATION:

NAME: Brown, Scott A.  
REGISTRATION NUMBER: 32,724  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617) 498-8224  
TELEFAX: (617) 876-5851

## INFORMATION FOR SEQ ID NO: 48:

SEQUENCE CHARACTERISTICS:  
LENGTH: 542 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear

MOLECULE TYPE: cDNA  
SEQUENCE DESCRIPTION: SEQ ID NO: 48:

US-10-040-916-48

## Query Match

Best Local Similarity 72.9%; Score 146.6; DB 12; Length 542;  
Matches 149; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

OY 39 AACAACTTCTTAAACATTACAGGAACCTATCCCTGAGAGAGGAGGAAAGAACTATTC 98  
|||||  
DB 1 AACAACTTCTTAAACATTACAGGAACCTATCCCTGAGAGAGGAGGAGGAAAGAACTATTC 60  
OY 99 ACCCTGTGACATGTGATAGTCACAGTCCCTCCCTATTCCTCCATCCCTAGATACAT 158  
|||||  
DB 61 ACCCTGTGACATGTGATAGTCACAGTCCCTCCCTATTCCTCCATCCCTAGATACAT 120  
OY 159 CCTGGAGAGACCTTACCCAGTCATTTTATCTACCCCACTGC 201  
|||||  
DB 121 CMTGGAGAGACCTTACCCAGTCATTTTATCTACCCCACTGC 163

## RESULT 4

US-08-979-847-112  
Sequence 112, Application US/08979847  
Publication No. US20030039664A1  
GENERAL INFORMATION:

APPLICANT: PERRON, HEVE  
APPLICANT: BESEME, FREDERIC  
APPLICANT: BEDIN, FREDERIC  
APPLICANT: PARANHOS-BACCALA, GLAUCIA  
APPLICANT: KOMURIAN-PRADEL, FLORENCE  
APPLICANT: JOLIVET-REYNAUD, COLETTE  
APPLICANT: MANDRAND, BERNARD  
APPLICANT: GARSON, JEREMY  
APPLICANT: TURK, PHILIP  
TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS  
TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYL  
NUMBER OF SEQUENCES: 210

```

CORRESPONDENCE ADDRESS:
ADDRESSEE: OLIFF & BERRIDGE, PLC
STREET: P. O. BOX 19928
CITY: ALEXANDRIA
STATE: VA
COUNTRY: USA
ZIP: 22320
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/979,847
FILING DATE: 26-NOV-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: BERRIDGE, WILLIAM P.
REGISTRATION NUMBER: 30,024
REFERENCE/DOCKET NUMBER: MPB 39046A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 703-836-6400
TELEFAX: 703-836-2787
INFORMATION FOR SEQ ID NO: 112:
SEQUENCE CHARACTERISTICS:
LENGTH: 758 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-979-847-112

Query Match      50.1%; Score 100.8; DB 7; Length 758;
Best Local Similarity 93.8%; Pred. No. 2.8e-23;
Matches 105; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 CAATGATGTCAGACATACCTCTTAGCCAAATATCAACAGTTCTTAACATTCACA 60
    |||||||
Db 647 CAATGATGTCAGACATACCTCTTAGCCAAATATCAACAGTTCTTAACATTCACA 706
QY 61 AGAAGCTATCCCTGAGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 112
    |||||||
Db 707 GGAAGCTGTCGCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 758

RESULT 5
US-09-864-761-20979/c
Sequence 20979, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aeomica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
PRIOR FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263,6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
```

```

PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
SEQ ID NO 20979
LENGTH: 200
TYPE: DNA
ORGANISM: Homo sapiens.
FEATURE:
OTHER INFORMATION: MAP TO AL034410.8
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 6.2
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 1.5
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.2
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.5
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.4
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 2.2
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.7
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.7
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.6
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.9
OTHER INFORMATION: NY HIT: g11419444, EVALUATE 5.00e-78
OTHER INFORMATION: EST HUMAN HIT: N76991.1, EVALUATE 1.00e-63
OTHER INFORMATION: SWISSPROT HIT: P03360, EVALUATE 2.00e-06
US-09-864-761-20979

Query Match      23.1%; Score 46.4; DB 10; Length 200;
Best Local Similarity 97.9%; Pred. No. 1.2e-05;
Matches 47; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 153 ATACATCTGTGGAGAGACCTTACCCAGTCATTTATCTACCCCACTG 200
    |||||||
Db 200 ATACATCTGTGGAGAGACCTTACCCAGTCATTTATCTACCCCACTG 153

RESULT 6
US-09-764-853-198/c
Sequence 198, Application US/09764853
Patent No. US20020090672A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: Puz06
CURRENT APPLICATION NUMBER: US/09/764,853
PRIOR FILING DATE: 2001-01-17
PRIOR APPLICATION DATA REMOVED - consult PALM or file wrapper
NUMBER OF SEQ ID NOS: 939
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 198
LENGTH: 2482
TYPE: DNA
ORGANISM: Homo sapiens
US-09-764-853-198
```

Query Match 16.6%; Score 33.4; DB 10; Length 2482;  
Best Local Similarity 54.5%; Pred. No. 0.7;  
Matches 67; Conservative 0; Mismatches 56; Indels 0; Gaps 0;

OY 31 CAAATATCAACAAGTTCTTAATAACATTACAGAACTATCCCTGAGAGGAAAGCA 90  
DB 595 CTTAAATGCAAACTTTTAACTATAGAAACCAATCTGCAACATATATAAG 536  
OY 91 ACTATTCACCCCTTGACATGATTAAGTCAAGTCCCTTCCTTAATTCCCATCCCT 150  
DB 535 GGTATTACATCATGATGAAAGTGCATTATCTGAAATCCAGGCTAATTCAATTCGA 476  
OY 151 AGA 153  
DB 475 AAA 473

## RESULT 7

US-09-767-088A-2  
; Sequence 2, Application US/09767088A  
; Patent No. US20020010947A1  
; GENERAL INFORMATION:  
; APPLICANT: Gurney, Mark E.  
; TITLE OF INVENTION: Transgenic Mouse Model Of Human Neurodegenerative Disease  
; FILE REFERENCE: PHRM0303  
; CURRENT APPLICATION NUMBER: US/09/767,088A  
; CURRENT FILING DATE: 2001-01-22  
; PRIOR APPLICATION NUMBER: 60/177,319  
; PRIOR FILING DATE: 2000-01-21  
; NUMBER OF SEQ ID NOS: 15  
; SOFTWARE: Patentin version 3.0  
; SEQ ID NO 2  
; LENGTH: 9990  
; TYPE: DNA  
; ORGANISM: Mus musculus  
US-09-767-088A-2

Query Match 16.3%; Score 32.8; DB 10; Length 9990;  
Best Local Similarity 59.8%; Pred. No. 2;  
Matches 55; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

OY 88 ACAATATTCACCCCTTGACATGATTAAGTCAAGTCCCTTCCTTAATTCGCCATC 147  
DB 7772 ACCAGTAAACAATCTGGAGGTATTATTAGTCGTGCTGACCCGAAACCCCACT 7831  
OY 148 CCTAGATACATCCTGGAGAGACCTTACCAG 179  
DB 7832 CCTGGCAATTTACTGGAGAGAAACAAG 7863

## RESULT 8

US-09-767-088A-15  
; Sequence 15, Application US/09767088A  
; Patent No. US20020010947A1  
; GENERAL INFORMATION:  
; APPLICANT: Gurney, Mark E.  
; APPLICANT: Abraham, Irene  
; TITLE OF INVENTION: Transgenic Mouse Model Of Human Neurodegenerative Disease  
; FILE REFERENCE: PHRM0303  
; CURRENT APPLICATION NUMBER: US/09/767,088A  
; CURRENT FILING DATE: 2001-01-22  
; PRIOR APPLICATION NUMBER: 60/177,319  
; PRIOR FILING DATE: 2000-01-21  
; NUMBER OF SEQ ID NOS: 15  
; SOFTWARE: Patentin version 3.0  
; SEQ ID NO 15  
; LENGTH: 9990  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; NAME/KEY: misc\_feature  
; OTHER INFORMATION: PrP/tau transgene construct

US-09-767-088A-15

Query Match 16.3%; Score 32.8; DB 10; Length 9990;  
Best Local Similarity 59.8%; Pred. No. 2;  
Matches 55; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

OY 88 ACAATATTCACCCCTTGACATGATTAAGTCAAGTCCCTTCCTTAATTCGCCATC 147  
DB 7772 ACCAGTAAACAATCTGGAGGTATTATTAGTCGTGCTGACCCGAAACCCCACT 7831  
OY 148 CCTAGATACATCCTGGAGAGACCTTACCAG 179  
DB 7832 CCTGGCAATTTACTGGAGAGAAACAAG 7863

## RESULT 9

US-09-764-868-1482  
; Sequence 1482, Application US/09764868  
; Patent No. US2002016871A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PT232  
; CURRENT APPLICATION NUMBER: US/09/764,868  
; CURRENT FILING DATE: 2001-01-17  
; Prior application data removed - refer to PAM or file wrapper  
; NUMBER OF SEQ ID NOS: 1510  
; SOFTWARE: Patentin Ver. 2.0  
; SEQ ID NO 1482  
; LENGTH: 10098  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-764-868-1482

Query Match 16.2%; Score 32.6; DB 9; Length 10098;  
Best Local Similarity 60.9%; Pred. No. 2.3;  
Matches 53; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

OY 69 ATCCCTGAGAAAGGAAAGAACTATTCACCCCTTGACATGATTAAGTCAAGTCC 128  
DB 1474 AACCCTGAGATTTGGGAAACTATTCATCTTTTTCACCTGATGATCTGCTCAT 1533  
OY 129 TTCCTCTAATTCCTCCATCCCTAGATA 155  
DB 1534 TTTCTCTGTTTCACCAACCTGAATA 1560

## RESULT 10

US-09-991-936-1452  
; Sequence 1452, Application US/09991936  
; Publication No. US20030073827A1  
; GENERAL INFORMATION:  
; APPLICANT: Brandt, Kevin S.  
; APPLICANT: Gaines, Patrick J.  
; APPLICANT: Slinchcomb, Dan T.  
; APPLICANT: Wisniewski, Nancy  
; TITLE OF INVENTION: NUCLEIC ACID MOLECULES, PROTEINS AND USES THEREOF  
; FILE REFERENCE: FC-6-C1  
; CURRENT APPLICATION NUMBER: US/09/991,936  
; CURRENT FILING DATE: 2001-11-21  
; PRIOR APPLICATION NUMBER: US/09/543,668  
; PRIOR FILING DATE: 2000-04-07  
; PRIOR APPLICATION NUMBER: 60/128,704  
; PRIOR FILING DATE: 1999-04-09  
; NUMBER OF SEQ ID NOS: 1959  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 1452  
; LENGTH: 550  
; TYPE: DNA  
; ORGANISM: Ctenocephalides felis  
US-09-991-936-1452



Query Match 15.6%; Score 31.4; DB 9; Length 550;  
Best Local Similarity 54.9%; Pred. No. 1.7;  
Matches 62; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

DB 28 AGCAATATCAACAATTCTTAAACATTCACAGAACCTATCCCTGAGAGGGGAAA 87  
224 AGCCAAATTTTACAAAGGCTCTTAATTTAAAGAAATATTAATGAGAACATTTTAA 283

DB 88 AGAATATTCACACCTGTGACATGATATTAAGCAATCCCTCCCTCTAAT 140  
284 TGAACATTTTAAACGTTAGCAATATTTGGAGTGTGCAAAATGCTTTAAAT 336

## RESULT 11

US-09-918-995-11931/c  
; Sequence 11931, Application US/09918995  
; Publication No. US2003073623A1  
; GENERAL INFORMATION:  
; APPLICANT: Hyseq, Inc.  
; TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
; FILE REFERENCE: 20411-756  
; CURRENT APPLICATION NUMBER: US/09/918,995  
; PRIOR FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: US/09/235,076  
; NUMBER OF SEQ ID NOS: 38054  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 11931  
; LENGTH: 509  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)...(509)  
; OTHER INFORMATION: n = A,T,C or G  
US-09-918-995-11931

Query Match 15.3%; Score 30.8; DB 9; Length 509;  
Best Local Similarity 50.7%; Pred. No. 2.5;  
Matches 74; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

DB 12 CAGACATCACCTCCTTACCAATATCAACAAGTTCTTAAACATTCACAGAACCTATC 71  
460 CAGTTCCTCTCCATCTTCTGATTTTGTAGATATAAATTTTAAACAAAGGATG 401  
DB 72 CCTGAGAGAGGAGAAAGAACTATTCACCCCTGTGACATGATTAAGTCAAGTCCCTTC 131  
400 TCCCTCTTCTAGTAGAGAACTATTCCTCAGCTCTGAGAGAGAGAAAAAAGATTC 341

DB 132 CCTCTAATTCCTCCATCCCTAGATACA 157  
340 CCATTTCTCTCCTACCTCAAAAACA 315

## RESULT 12

US-09-822-846-42/c  
; Sequence 42, Application US/09822846  
; Publication No. US20030027139A1  
; GENERAL INFORMATION:  
; APPLICANT: Jacobs, Kenneth  
; APPLICANT: McCoy, John M.  
; APPLICANT: Lavallee, Edward R.  
; APPLICANT: Collins-Racie, Lisa A.  
; APPLICANT: Evans, Cheryl  
; APPLICANT: Merberg, David  
; APPLICANT: Treacy, Maurice  
; APPLICANT: Agostino, Michael J.  
; APPLICANT: Steininger II, Robert J.  
; APPLICANT: Spaulding, Vikki  
; APPLICANT: Wong, Gordon G.  
; APPLICANT: Clark, Hilary

APPLICANT: Fechtel, Kim  
; APPLICANT: Howes, Steven H.  
; APPLICANT: Resnick, Richard J.  
; APPLICANT: Gulukota, Kamalakara  
; APPLICANT: Graham, James R.  
; APPLICANT: Genetics Institute, Inc.  
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS  
; FILE REFERENCE: GIN 6400  
; CURRENT APPLICATION NUMBER: US/09/822,846  
; PRIOR FILING DATE: 2001-03-29  
; PRIOR APPLICATION NUMBER: 60/195,605  
; NUMBER OF SEQ ID NOS: 629  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 42  
; LENGTH: 2456  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-822-846-42

Query Match 15.3%; Score 30.8; DB 9; Length 2456;  
Best Local Similarity 50.7%; Pred. No. 5;  
Matches 74; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

DB 13 AGACATCACCTCCTTACCAATATCAACAAGTTCTTAAACATTCACAGAACCTATCC 72  
634 AATCTCACTTTTCCCAACACACATTAATTTATAGAAACACTAGATAGACTCTCC 575  
DB 73 CTGAGAGAGGAGAAAGAACTATTCACCCCTGTGACATGATTAAGTCAAGTCCCTTC 132  
574 ATGACAATGACAGTAGAGAAACATCTCAAAACCTCTCTAGTGTGTCAGGCTCAAG 515

DB 133 CTCCTAATTCCTCCATCCCTAGATACAT 158  
514 ATTGCTTACCAATTTGTCTAAT 489

## RESULT 13

US-10-114-893-122/c  
; Sequence 122, Application US/10114893  
; Publication No. US20020193567A1  
; GENERAL INFORMATION:  
; APPLICANT: McCoy, John M.  
; APPLICANT: Lavallee, Edward R.  
; APPLICANT: Collins-Racie, Lisa A.  
; APPLICANT: Evans, Cheryl  
; APPLICANT: Merberg, David  
; APPLICANT: Treacy, Maurice  
; APPLICANT: Bowman, Michael R.  
; APPLICANT: Spaulding, Vikki  
; APPLICANT: Carlin-Duckett, McKeough  
; APPLICANT: Kelleher, Kerry S.  
; APPLICANT: Genetics Institute, Inc.  
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM  
; FILE REFERENCE: GI 6000-10A  
; CURRENT APPLICATION NUMBER: US/10/114,893  
; PRIOR FILING DATE: 2002-04-02  
; EARLIER APPLICATION NUMBER: 09/413,232  
; NUMBER OF SEQ ID NOS: 321  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 122  
; LENGTH: 2496  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-114-893-122

Query Match 15.3%; Score 30.8; DB 9; Length 2496;  
Best Local Similarity 50.7%; Pred. No. 5;  
Matches 74; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

DB 13 AGACATCACCTCCTTACCAATATCAACAAGTTCTTAAACATTCACAGAACCTATCC 72

DB 634 AATCTCACTTTTCCCAACACACTAAATTTATAGAAACACTAGATAGCTCTCC 575  
QY 73 CTGAGAGAGGGGAAAGAACTATTCCACCTGTGACATGTTATGTCAAGTCCCTTC 132  
DB 574 ATGACAAATGACAGTGAAGAAACATCTCAAAAGCTCTCTAGTTGTGTCCAGGGCTCAAG 515  
QY 133 CTCTAATTCCTCCATTCCTAGATACAT 158  
DB 514 ATTGCTTCACCAAAATTTGCTAAAT 489

## RESULT 14

US-10-092-154-1965  
; Sequence 1965, Application US/10092154  
; Publication No. US20030054375A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PC009C1  
; CURRENT APPLICATION NUMBER: US/10/092,154  
; CURRENT FILING DATE: 2002-03-07  
; NUMBER OF SEQ ID NOS: 2003  
; Prior Application removed - See File Wrapper or Palm  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO: 1965  
; LENGTH: 26197  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-092-154-1965

Query Match 15.3%; Score 30.8; DB 9; Length 26197;  
Best Local Similarity 50.7%; Pred. No. 14;  
Matches 74; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

QY 13 AGACATCACCCTCTGTCCTCAATATACAGATCTTAAACATTCAGAGAACTATCC 72  
DB 3270 AATCTCACTTTTCCCAACACACTAAATTTATAGAAACACTAGATAGCTCTCC 3329  
QY 73 CTGAGAGAGGGGAAAGAACTATTCCACCTGTGACATGTTATGTCAAGTCCCTTC 132  
DB 3330 ATGACAAATGACAGTGAAGAAACATCTCAAAAGCTCTCTAGTTGTGTCCAGGGCTCAAG 3389  
QY 133 CTCTAATTCCTCCATTCCTAGATACAT 158  
DB 3390 ATTGCTTCACCAAAATTTGCTAAAT 3415

## RESULT 15

US-09-764-847-1965  
; Sequence 1965, Application US/09764847  
; Patent No. US20020132767A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
; FILE REFERENCE: PC009  
; CURRENT APPLICATION NUMBER: US/09/764,847  
; CURRENT FILING DATE: 2001-01-17  
; Prior application data removed - consult PALM or file wrapper  
; NUMBER OF SEQ ID NOS: 2003  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO: 1965  
; LENGTH: 26197  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-764-847-1965

Query Match 15.3%; Score 30.8; DB 10; Length 26197;  
Best Local Similarity 50.7%; Pred. No. 14;  
Matches 74; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

QY 13 AGACATCACCCTCTGTCCTCAATATACAGATCTTAAACATTCAGAGAACTATCC 72  
DB 3270 AATCTCACTTTTCCCAACACACTAAATTTATAGAAACACTAGATAGCTCTCC 3329  
QY 73 CTGAGAGAGGGGAAAGAACTATTCCACCTGTGACATGTTATGTCAAGTCCCTTC 132  
DB 3330 ATGACAAATGACAGTGAAGAAACATCTCAAAAGCTCTCTAGTTGTGTCCAGGGCTCAAG 3389  
QY 133 CTCTAATTCCTCCATTCCTAGATACAT 158  
DB 3390 ATTGCTTCACCAAAATTTGCTAAAT 3415

DB 3270 AATCTCACTTTTCCCAACACACTAAATTTATAGAAACACTAGATAGCTCTCC 3329  
QY 73 CTGAGAGAGGGGAAAGAACTATTCCACCTGTGACATGTTATGTCAAGTCCCTTC 132  
DB 3330 ATGACAAATGACAGTGAAGAAACATCTCAAAAGCTCTCTAGTTGTGTCCAGGGCTCAAG 3389  
QY 133 CTCTAATTCCTCCATTCCTAGATACAT 158  
DB 3390 ATTGCTTCACCAAAATTTGCTAAAT 3415

Search completed: May 2, 2003, 18:18:23  
Job time : 86.3333 secs

GenCore version 5.1.5  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 2, 2003, 12:54:45 ; Search time 689 Seconds

(without alignments)  
8490.078 Million cell updates/sec

Title: US-09-719-554-3\_COPY\_8500\_8700

Perfect score: 201  
Sequence: 1 aactcagcagcaagaataaa.....tatttttgcgtgtracc 201

Scoring table:

IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl: \*  
1: gb\_da: \*  
2: gb\_htg: \*  
3: gb\_in: \*  
4: gb\_ov: \*  
5: gb\_ov: \*  
6: gb\_pat: \*  
7: gb\_ph: \*  
8: gb\_pl: \*  
9: gb\_pr: \*  
10: gb\_ro: \*  
11: gb\_sts: \*  
12: gb\_sy: \*  
13: gb\_un: \*  
14: gb\_vl: \*  
15: em\_ba: \*  
16: em\_fun: \*  
17: em\_hum: \*  
18: em\_in: \*  
19: em\_mu: \*  
20: em\_om: \*  
21: em\_or: \*  
22: em\_ov: \*  
23: em\_ov: \*  
24: em\_pat: \*  
25: em\_pl: \*  
26: em\_ro: \*  
27: em\_sts: \*  
28: em\_un: \*  
29: em\_vl: \*  
30: em\_htg\_hum: \*  
31: em\_htg\_inv: \*  
32: em\_htg\_other: \*  
33: em\_htg\_mus: \*  
34: em\_htg\_pln: \*  
35: em\_htg\_rnd: \*  
36: em\_htg\_mam: \*  
37: em\_htg\_vrl: \*  
38: em\_sy: \*  
39: em\_htgo\_hum: \*  
40: em\_htgo\_mus: \*  
41: em\_htgo\_other: \*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	201	100.0	540	9 AB051004	AB051004 Homo sapi
2	201	100.0	540	9 AB079501	AB079501 Homo sapi
3	201	100.0	541	9 AB079503	AB079503 Homo sapi
4	201	100.0	541	9 AB079504	AB079504 Homo sapi
5	201	100.0	541	9 AB079509	AB079509 Homo sapi
6	201	100.0	542	9 AB079502	AB079502 Homo sapi
7	201	100.0	543	9 AB054089	AB054089 Homo sapi
8	201	100.0	543	9 AB055416	AB055416 Homo sapi
9	201	100.0	1617	9 AF156963	AF156963 Homo sapi
10	201	100.0	1624	9 AF506835	AF506835 Homo sapi
11	201	100.0	1860	9 AF513360	AF513360 Homo sapi
12	201	100.0	1948	6 AX000960	AX000960 Sequence
13	201	100.0	1948	6 AX027474	AX027474 Sequence
14	201	100.0	1948	9 AF072505	AF072505 Homo sapi
15	201	100.0	2055	6 AX007999	AX007999 Sequence
16	201	100.0	2599	6 AX007978	AX007978 Sequence
17	201	100.0	2781	6 AX092223	AX092223 Sequence
18	201	100.0	2781	6 AF072506	AF072506 Homo sapi
19	201	100.0	2930	6 AX355872	AX355872 Sequence
20	201	100.0	2930	9 AF208161	AF208161 Homo sapi
21	201	100.0	2946	6 AR177269	AR177269 Sequence
22	201	100.0	10499	6 AX007980	AX007980 Sequence
23	201	100.0	56093	6 AX329572	AX329572 Sequence
24	201	100.0	56093	9 HSAC000064	HSAC000064 Human BAC
25	201	100.0	149194	9 AC007566	AC007566 Homo sapi
26	200.6	99.8	7582	6 AX000966	AX000966 Sequence
27	200.6	99.8	7582	6 AX027480	AX027480 Sequence
28	199.4	99.2	2782	6 AX000962	AX000962 Sequence
29	199.4	99.2	2782	6 AX027476	AX027476 Sequence
30	199.4	99.2	2782	6 AX133396	AX133396 Sequence
31	194.6	96.8	540	9 AB055412	AB055412 Homo sapi
32	193	96.0	540	9 AB051010	AB051010 Homo sapi
33	193	96.0	1481	6 AX001027	AX001027 Sequence
34	193	96.0	1629	6 AX036992	AX036992 Sequence
35	193	96.0	1629	14 AF331500	AF331500 Multiple
36	193	96.0	1799	6 AX007982	AX007982 Sequence
37	193	96.0	1932	14 AF127228	AF127228 Multiple
38	193	96.0	46340	6 AX008043	AX008043 Sequence
39	193	96.0	128485	2 AL358814	AL358814 Homo sapi
40	193	96.0	162579	9 AL390039	AL390039 Human DNA
41	193	96.0	172918	2 AC023366	AC023366 Homo sapi
42	193	96.0	177163	9 HS134E15	AL022067 Human DNA
43	191.4	95.2	539	9 AB054086	AB054086 Homo sapi
44	191.4	95.2	540	9 AB051003	AB051003 Homo sapi
45	191.4	95.2	540	9 AB051007	AB051007 Homo sapi

#### ALIGNMENTS

RESULT 1  
AB051004 540 bp DNA linear PRI 18-MAY-2001  
LOCUS  
DEFINITION Homo sapiens human endogenous retrovirus W gene for envelope protein, partial cds, isolate:W-7-1.  
ACCESSION AB051004  
VERSION AB051004.1 GI:14020829  
KEYWORDS  
SOURCE  
ORGANISM Homo sapiens (isolate:W-7-1) DNA.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
AUTHORS Kim/H.S. and Lee/W.H.  
TITLE Human endogenous retrovirus HERV-W family: chromosomal

Pred. No. is the number of results predicted by chance to have a

JOURNAL Localization, identification, and phylogeny  
MEDLINE AIDS Res. Hum. Retroviruses 17 (7), 643-648 (2001)  
REFERENCE 2 (bases 1 to 540)

AUTHORS Kim, H.  
TITLE Direct Submission  
JOURNAL Submitted (10-NOV-2000) Heui-Seo Kim, Pusan National University,  
Biology, Keumjeong Ku, Changjeon Dong, Pusan 609-735, Korea  
(E-mail: khs307@hyowon.cc.pusan.ac.kr, URL: http://mpl.biology.or.kr,  
Tel: 82515102259, Fax: 825151812962)

FEATURES Location/Qualifiers

source  
1..540  
/organism="Homo sapiens"  
/isolate="W-7-1"  
/db\_xref="taxon:9606"  
/chromosome="7"  
1..540  
/gene="env"  
<1..>540  
/gene="env"  
/codon\_start=1  
/product="envelope protein"  
/protein\_id="BAB47558.1"  
/db\_xref="GI:14020830"  
/translation="NNEPHEINTSVLVGPIVSNLEIHTHTNLGCVKRSNTYTTNSO  
CIRWTPPTQIVCLPSGIFVCGTSAYRGLNSSSMGLSPVPMITYTODLXY  
VTSRPNKRVPLPPIVAGVIGALGTGIGTITSTOPIYKLSQELNGMERYADSLV  
TLQDNLSTAAVYLVIONRRAL"  
repeat\_region  
1..540  
/rpt\_family="HERV-W"  
BASE COUNT 156 a 146 c 89 g 149 t  
ORIGIN

Query Match 100.0%; Score 201; DB 9; Length 540;  
Best Local Similarity 100.0%; Pred. No. 1.1e-48;  
Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AACTCAGCAGAGAAATAAACACCACTCCGTTTGTAGTAGGACCTCTGTTTCCAAATCG 60  
|||||  
DB 4 AACTTCAGCAGAGAAATAAACACCACTCCGTTTGTAGTAGGACCTCTGTTTCCAAATCG 63  
|||||  
QY 61 GAATATACCATCTCAACCTCAACCTGCTGTAATTTAGCAATCTACATACAGAAC 120  
|||||  
DB 64 GAATATACCATCTCAACCTCAACCTGCTGTAATTTAGCAATCTACATACAGAAC 123  
|||||  
QY 121 AACTCCCAATGCATCAGGTGGGTAACTCCTCCACACAATACTGCTACCCCTCAGGA 180  
|||||  
DB 124 AACTCCCAATGCATCAGGTGGGTAACTCCTCCACACAATACTGCTACCCCTCAGGA 183  
|||||  
QY 181 ATATTTTGTCTGTGTGTACC 201  
|||||  
DB 184 ATATTTTGTCTGTGTGTACC 204  
|||||

RESULT 2  
AB079501 540 bp mRNA linear PRI 06-AUG-2002  
LOCUS Homo sapiens mRNA, repeat sequence HERV-W, isolate:HB-1.  
DEFINITION AB079501  
ACCESSION AB079501  
VERSION AB079501.1 GI:22122075  
KEYWORDS  
SOURCE Homo sapiens (isolate:HB-1) brain CDNA to mRNA.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1  
AUTHORS Shin, K., Yi, J. and Kim, H.  
TITLE Isolation and Phylogeny of HERV-W env Fragments in Human Brain CDNA  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 540)  
AUTHORS Shin, K., Yi, J. and Kim, H.  
TITLE Direct Submission

JOURNAL Submitted (05-FEB-2002) Heui-Seo Kim, Pusan National University,  
Biology, 30 Changjeon-dong, Keumjeong-ku, Pusan 609-735, Republic  
of Korea (E-mail: khs307@hyowon.cc.pusan.ac.kr, Tel: 82515103346,  
Fax: 825151812962)

FEATURES Location/Qualifiers

source  
1..540  
/organism="Homo sapiens"  
/isolate="HB-1"  
/db\_xref="taxon:9606"  
/tissue\_type="brain"  
repeat\_region  
1..540  
/rpt\_family="HERV-W"  
BASE COUNT 156 a 146 c 89 g 149 t  
ORIGIN

Query Match 100.0%; Score 201; DB 9; Length 540;  
Best Local Similarity 100.0%; Pred. No. 1.1e-48;  
Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AACTCAGCAGAGAAATAAACACCACTCCGTTTGTAGTAGGACCTCTGTTTCCAAATCG 60  
|||||  
DB 4 AACTTCAGCAGAGAAATAAACACCACTCCGTTTGTAGTAGGACCTCTGTTTCCAAATCG 63  
|||||  
QY 61 GAATATACCATCTCAACCTCAACCTGCTGTAATTTAGCAATCTACATACAGAAC 120  
|||||  
DB 64 GAATATACCATCTCAACCTCAACCTGCTGTAATTTAGCAATCTACATACAGAAC 123  
|||||  
QY 121 AACTCCCAATGCATCAGGTGGGTAACTCCTCCACACAATACTGCTACCCCTCAGGA 180  
|||||  
DB 124 AACTCCCAATGCATCAGGTGGGTAACTCCTCCACACAATACTGCTACCCCTCAGGA 183  
|||||  
QY 181 ATATTTTGTCTGTGTGTACC 201  
|||||  
DB 184 ATATTTTGTCTGTGTGTACC 204  
|||||

RESULT 3  
AB079503 541 bp mRNA linear PRI 06-AUG-2002  
LOCUS Homo sapiens mRNA, repeat sequence HERV-W, isolate:HB-3.  
DEFINITION AB079503  
ACCESSION AB079503  
VERSION AB079503.1 GI:22122077  
KEYWORDS  
SOURCE Homo sapiens (isolate:HB-3) brain CDNA to mRNA.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1  
AUTHORS Shin, K., Yi, J. and Kim, H.  
TITLE Isolation and Phylogeny of HERV-W env Fragments in Human Brain CDNA  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 541)  
AUTHORS Shin, K., Yi, J. and Kim, H.  
TITLE Direct Submission  
JOURNAL Submitted (05-FEB-2002) Heui-Seo Kim, Pusan National University,  
Biology, 30 Changjeon-dong, Keumjeong-ku, Pusan 609-735, Republic  
of Korea (E-mail: khs307@hyowon.cc.pusan.ac.kr, Tel: 82515103346,  
Fax: 825151812962)

FEATURES Location/Qualifiers

source  
1..541  
/organism="Homo sapiens"  
/isolate="HB-3"  
/db\_xref="taxon:9606"  
/tissue\_type="brain"  
repeat\_region  
1..541  
/rpt\_family="HERV-W"  
BASE COUNT 154 a 146 c 92 g 149 t  
ORIGIN

Query Match 100.0%; Score 201; DB 9; Length 541;  
Best Local Similarity 100.0%; Pred. No. 1.1e-48;  
Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AACTTCACGACGAAATAACACACCTCGTTTAGTAGGACCTGTTTCCAAATCG 60  
LOCUS AB079504 541 bp mRNA linear PRI 06-AUG-2002  
DEFINITION Homo sapiens mRNA, repeat sequence HERV-W, isolate:HB-4.  
ACCESSION AB079504  
VERSION AB079504.1 GI:22122078  
KEYWORDS  
SOURCE Homo sapiens (isolate:HB-4) brain cDNA to mRNA.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1  
AUTHORS Shin, K., Yi, J., and Kim, H.  
TITLE Isolation and Phylogeny of HERV-W env Fragments in Human Brain cDNA library  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 541)  
AUTHORS Shin, K., Yi, J., and Kim, H.  
TITLE Direct Submission  
JOURNAL Submitted (05-FEB-2002) Heui-Soo Kim, Pusan National University, Biology; 30 Changjeon-dong Keumjeoung-ku, Pusan 609-735, Republic of Korea (E-mail: khs307@hyowon.cc.pusan.ac.kr, Tel: 82515103346, Fax: 82515812962)  
FEATURES  
source location/Qualifiers  
1. 541  
/organism="Homo sapiens"  
/isolate="HB-4"  
/db\_xref="taxon:9606"  
/tissue\_type="brain"  
repeat\_region 1. 541  
/rpl\_family="HERV-W"  
BASE COUNT 155 a 144 c 90 g 152 t  
ORIGIN  
Query Match 100.0%; Score 201; DB 9; Length 541;  
Best Local Similarity 100.0%; Pred. No. 1,1e-48;  
Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 AACTTCACGACGAAATAACACACCTCGTTTAGTAGGACCTGTTTCCAAATCG 60  
Db 4 AACTTCACGACGAAATAACACACCTCGTTTAGTAGGACCTGTTTCCAAATCG 63  
OY 61 GAATAAACCCATACCTCAACCTGCTGTAAATTTAGCAATACATACACCAACC 120  
Db 64 GAATAAACCCATACCTCAACCTGCTGTAAATTTAGCAATACATACACCAACC 123  
OY 121 AACTCCCATGATGAGTGGGTAACTCTCCACACAAATAGTCTGCTACCTCAGGA 180  
Db 124 AACTCCCATGATGAGTGGGTAACTCTCCACACAAATAGTCTGCTACCTCAGGA 183  
OY 181 ATATTTTGTCTGTGTAC 201  
Db 184 ATATTTTGTCTGTGTAC 204

RESULT 5  
AB079509 541 bp mRNA linear PRI 06-AUG-2002  
LOCUS AB079509  
DEFINITION Homo sapiens mRNA, repeat sequence HERV-W, isolate:HB-9.  
ACCESSION AB079509  
VERSION AB079509.1 GI:22122083  
KEYWORDS  
SOURCE Homo sapiens (isolate:HB-9) brain cDNA to mRNA.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1  
AUTHORS Shin, K., Yi, J., and Kim, H.  
TITLE Isolation and Phylogeny of HERV-W env Fragments in Human Brain cDNA library  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 541)  
AUTHORS Shin, K., Yi, J., and Kim, H.  
TITLE Direct Submission  
JOURNAL Submitted (05-FEB-2002) Heui-Soo Kim, Pusan National University, Biology; 30 Changjeon-dong Keumjeoung-ku, Pusan 609-735, Republic of Korea (E-mail: khs307@hyowon.cc.pusan.ac.kr, Tel: 82515103346, Fax: 82515812962)  
FEATURES  
source location/Qualifiers  
1. 541  
/organism="Homo sapiens"  
/isolate="HB-9"  
/db\_xref="taxon:9606"  
/tissue\_type="brain"  
repeat\_region 1. 541  
/rpl\_family="HERV-W"  
BASE COUNT 156 a 146 c 89 g 150 t  
ORIGIN  
Query Match 100.0%; Score 201; DB 9; Length 541;  
Best Local Similarity 100.0%; Pred. No. 1,1e-48;  
Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 AACTTCACGACGAAATAACACACCTCGTTTAGTAGGACCTGTTTCCAAATCG 60  
Db 4 AACTTCACGACGAAATAACACACCTCGTTTAGTAGGACCTGTTTCCAAATCG 63  
OY 61 GAATAAACCCATACCTCAACCTGCTGTAAATTTAGCAATACATACACCAACC 120  
Db 64 GAATAAACCCATACCTCAACCTGCTGTAAATTTAGCAATACATACACCAACC 123  
OY 121 AACTCCCATGATGAGTGGGTAACTCTCCACACAAATAGTCTGCTACCTCAGGA 180  
Db 124 AACTCCCATGATGAGTGGGTAACTCTCCACACAAATAGTCTGCTACCTCAGGA 183  
OY 181 ATATTTTGTCTGTGTAC 201  
Db 184 ATATTTTGTCTGTGTAC 204  
RESULT 6  
AB079502 542 bp mRNA linear PRI 06-AUG-2002  
LOCUS AB079502  
DEFINITION Homo sapiens mRNA, repeat sequence HERV-W, isolate:HB-2.  
ACCESSION AB079502  
VERSION AB079502.1 GI:22122076  
KEYWORDS  
SOURCE Homo sapiens (isolate:HB-2) brain cDNA to mRNA.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1  
AUTHORS Shin, K., Yi, J., and Kim, H.  
TITLE Isolation and Phylogeny of HERV-W env Fragments in Human Brain cDNA library  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 542)  
AUTHORS Shin, K., Yi, J., and Kim, H.

```

TITLE      Direct Submission
JOURNAL    Submitted (05-FEB-2002) Heui-Soo Kim, Pusan National University,
            Biology; 30 Changjeon-dong, Keunjeoung-ku, Pusan 609-735, Republic
            of Korea (E-mail: khs307ehyowon.cc.pusan.ac.kr, Tel:82515103346,
            Fax:82515812962)
FEATURES   Location/Qualifiers
            source
              1..542
                /organism="Homo sapiens"
                /isolate="HB-2"
                /db_xref="taxon:9606"
                /tissue_type="brain"
            repeat_region
              1..542
                /rpt_family="HERV-W"
BASE COUNT 157 a 145 c 91 g 149 t
ORIGIN
Query Match      100.0%; Score 201; DB 9; Length 542;
Best Local Similarity 100.0%; Pred. No. 1.le-48;
Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AACTGAGCAGCAAAATAAACACCACTCCGTTTGTAGTAGACCTCTGTTCCAACTCG 60
    |||||||
DB 4 AACTTGAGCAGCAAAATAAACACCACTCCGTTTGTAGTAGACCTCTGTTCCAACTCG 63
OY 61 GAAATAACCCATACCTCAAACTCAGCTGTGTAAATTTAGCAATACATACACAAACC 120
    |||||||
DB 64 GAAATAACCCATACCTCAAACTCAGCTGTGTAAATTTAGCAATACATACACAAACC 123
OY 121 AACTCCCAATGATCAGTGGGTGTTACTCTCCACACAATAATAGTGCCTACCTCAGGA 180
    |||||||
DB 124 AACTCCCAATGATCAGTGGGTGTTACTCTCCACACAATAATAGTGCCTACCTCAGGA 183
OY 181 ATATTTTGTCTGTGTGTAC 201
    |||||||
DB 184 ATATTTTGTCTGTGTGTAC 204

RESULT 7
LOCUS      AB054089 543 bp DNA linear PRI 06-DEC-2001
DEFINITION Homo sapiens DNA, human endogenous retrovirus HERV-W family
ACCESSION  AB054089
VERSION     AB054089.1 GI:14780053
KEYWORDS
SOURCE      Homo sapiens (isolate:UO31-3) cell_line:UO-31 DNA.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE   1
            Yl,J.-M., Kim,H.-M., Lee,W.-H. and Kim,H.-S.
            Molecular cloning and phylogenetic analysis of new human endogenous
            retrovirus HERV-W family in cancer cells
            JOURNAL Curr. Microbiol. (2001) In press
            AUTHORS Lee,J. and Kim,H.-S.
            TITLE Direct Submission
            JOURNAL Submitted (15-JAN-2001) Joo-Mi Lee, Pusan National University,
            Biology; 30 Changjeon-dong, Keunjeoung-ku, Pusan 609-735, Republic
            of Korea (E-mail:jmlee76@hotmail.com, Tel:82515103346,
            Fax:82515812962)
FEATURES   Location/Qualifiers
            source
              1..543
                /organism="Homo sapiens"
                /isolate="UO31-3"
                /db_xref="taxon:9606"
                /cell_line="UO-31"
                /note="HERV-W env"
            repeat_region
              1..543
                /rpt_family="human endogenous retrovirus HERV-W"
BASE COUNT 156 a 146 c 89 g 152 t
ORIGIN
Query Match      100.0%; Score 201; DB 9; Length 543;
Best Local Similarity 100.0%; Pred. No. 1.le-48;
Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AACTGAGCAGCAAAATAAACACCACTCCGTTTGTAGTAGACCTCTGTTCCAACTCG 60
    |||||||
DB 4 AACTTGAGCAGCAAAATAAACACCACTCCGTTTGTAGTAGACCTCTGTTCCAACTCG 63
OY 61 GAAATAACCCATACCTCAAACTCAGCTGTGTAAATTTAGCAATACATACACAAACC 120
    |||||||
DB 64 GAAATAACCCATACCTCAAACTCAGCTGTGTAAATTTAGCAATACATACACAAACC 123
OY 121 AACTCCCAATGATCAGTGGGTGTTACTCTCCACACAATAATAGTGCCTACCTCAGGA 180
    |||||||
DB 124 AACTCCCAATGATCAGTGGGTGTTACTCTCCACACAATAATAGTGCCTACCTCAGGA 183

RESULT 8
LOCUS      AB055416 543 bp DNA linear PRI 07-DEC-2001
DEFINITION Homo sapiens DNA, human endogenous retrovirus HERV-W family
ACCESSION  AB055416
VERSION     AB055416.1 GI:15128498
KEYWORDS
SOURCE      Homo sapiens (isolate:RT4-2) cell_line:RT4 DNA.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE   1
            Yl,J.-M., Lee,W.-H., Kim,H.-M. and Kim,H.-S.
            Identification of new endogenous retroviral sequences belonging to
            the HERV-W family in human cancer cells
            JOURNAL Intervirology (2001) In press
            AUTHORS Lee,J. and Kim,H.
            TITLE Direct Submission
            JOURNAL Submitted (04-FEB-2001) Joo-Mi Lee, Pusan National University,
            Biology; 30 Changjeon-dong, Keunjeoung-ku, Pusan 609-735, Republic
            of Korea (E-mail:jmlee76@hotmail.com, Tel:82515103346,
            Fax:82515812962)
FEATURES   Location/Qualifiers
            source
              1..543
                /organism="Homo sapiens"
                /isolate="RT4-2"
                /db_xref="taxon:9606"
                /cell_line="RT4"
                /note="HERV-W env"
            repeat_region
              1..543
                /rpt_family="human endogenous retrovirus HERV-W"
BASE COUNT 156 a 146 c 89 g 152 t
ORIGIN
Query Match      100.0%; Score 201; DB 9; Length 543;
Best Local Similarity 100.0%; Pred. No. 1.le-48;
Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 AACTGAGCAGCAAAATAAACACCACTCCGTTTGTAGTAGACCTCTGTTCCAACTCG 60
    |||||||
DB 4 AACTTGAGCAGCAAAATAAACACCACTCCGTTTGTAGTAGACCTCTGTTCCAACTCG 63
OY 61 GAAATAACCCATACCTCAAACTCAGCTGTGTAAATTTAGCAATACATACACAAACC 120
    |||||||
DB 64 GAAATAACCCATACCTCAAACTCAGCTGTGTAAATTTAGCAATACATACACAAACC 123
OY 121 AACTCCCAATGATCAGTGGGTGTTACTCTCCACACAATAATAGTGCCTACCTCAGGA 180
    |||||||
DB 124 AACTCCCAATGATCAGTGGGTGTTACTCTCCACACAATAATAGTGCCTACCTCAGGA 183
```

QY 181 ATATTTTGTCTGTGTACC 201  
|||||  
Db 184 ATATTTTGTCTGTGTACC 204

RESULT 9  
AF156963 1617 bp DNA linear PRI 06-JUN-2000  
LOCUS Homo sapiens human endogenous retrovirus W envC7-1 envelope protein  
DEFINITION (env) gene, complete cds.  
ACCESSION AF156963  
VERSION AF156963.1 GI:8272467  
KEYWORDS  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 1617)  
Voisset,C., Bouton,O., Bedin,F., Duret,L., Mandrand,B., Mallet,F.  
and Paranhos-Baccala,G.  
Chromosomal distribution and coding capacity of the human  
endogenous retrovirus HERV-W family  
AIDS Res. Hum. Retroviruses 16 (8), 731-740 (2000)  
JOURNAL MEDLINE  
20284713  
PUBMED 10826480  
2 (bases 1 to 1617)  
Voisset,C., Bouton,O., Bedin,F., Duret,L., Mandrand,B., Mallet,F.  
and Paranhos-Baccala,G.  
TITLE Direct Submission  
JOURNAL Submitted (09-JUN-1999) UMRI03CNRS-biomerieux, ENS Lyon, 46 allée  
d'Italie, Lyon 69364, France  
FEATURES  
source  
1. 1617  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="envC7-1"  
/note="human/rodent monochromosomal NIGMS somatic hybrid  
panel #2 (Coriell Institute, Camden, NJ, USA); clone  
GM/NA10791 retaining human chromosome 7"  
1. 1617  
/gene="env"  
/gene="env"  
1. 1617  
/gene="env"  
/codon\_start=1  
/product="envelope protein"  
/protein\_id="AA074215.1"  
/db\_xref="GI:8272468"  
/translation="MALPYHIFLFTVLLPSFTLLAPPCRCMTSSSPYOEFLMMPOR  
GNIDAPYSRSLSGKPTPTATHTHPRNCYHSATLCMHANTHYGKMINPSCPGGLV  
TVCWTVFTQGMDSGGVODARREKHVEVLSQTRVHGTSPPYKGLDLSKLTETLT  
HTRLVSLFNTTLGLHEVSAQNPNCMCICPLNRPVYSIPVPOMNNFSEINTTSV  
LVGLVSNLEITHRSNLTGVKFSNTYTTNSOCIRWVTPPOIICLPSGIFVCGTSA  
YRCLNGSSSMCEFLSPVPMPTIYEDDLVYVSKPRNRPVILPVIAGVIGALG  
TGIGITTSQFYKLSQELNGEMERVADSLVTLQDOLNSIAAVILNRRALDILTAE  
RGGTCFLGEECCYYNOSGIVTKVKEIRRIORRAEELNRPWGLLSQMMWILP  
FLGLAAILILLFLGPCIENLVNFVSSRIEAVKLQMPKQSKTKTKYRRPLDRPAS  
RSDVNDIKGTPPEIISAQPLRNAGSS"  
repeat\_region  
1. 1617  
/rpt\_family="human endogenous retrovirus W"  
/rpt\_type="dispersed"  
1. 18  
/gene="env"  
1597. 1617  
/gene="env"  
BASE COUNT 441 a 452 c 297 g 427 t  
ORIGIN

Query Match 100.0%; Score 201; DB 9; Length 1617;  
Best Local Similarity 100.0%; Freq. No. 1e-48;  
Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AACTTCAGCAGCAAGAAATAACACACCTCCGTTTAGTAGAGACCTCTGTTCCATCTG 60

Db 622 AACTTCAGCAGCAAGAAATAACACACCTCCGTTTAGTAGAGACCTCTGTTCCATCTG 681  
|||||  
QY 61 GAATTAACCCATCTCTCAACCTCAGCTGTGTAATAATTAGCATATCTACCAACC 120  
|||||  
Db 682 GAATTAACCCATCTCTCAACCTCAGCTGTGTAATAATTAGCATATCTACCAACC 741  
|||||

QY 121 AACTCCCAATGATGAGGTGGGTAGCTCTCCACACAATAGTGTGCTACCCAGGA 180  
|||||  
Db 742 AACTCCCAATGATGAGGTGGGTAGCTCTCTCCACACAATAGTGTGCTACCCAGGA 801  
|||||

QY 181 ATATTTTGTCTGTGTACC 201  
|||||  
Db 802 ATATTTTGTCTGTGTACC 822

RESULT 10  
AF506835 1624 bp mRNA linear PRI 20-MAY-2002  
LOCUS Homo sapiens enverin mRNA, partial cds.  
DEFINITION AF506835  
ACCESSION AF506835  
VERSION AF506835.1 GI:20978307  
KEYWORDS  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 1624)  
Alliel,P.M., Perin,J.P., Pierig,R., Nusbaum,J.L., Menard,A. and  
Rieger,F.  
Endogenous retroviruses and multiple sclerosis. II. HERV-7q  
C. R. Acad. Sci. III, Sci. Vie 321 (10), 857-863 (1998)  
JOURNAL MEDLINE  
99052087  
PUBMED 9835022  
2 (bases 1 to 1624)  
Alliel,P.M., Perin,J.P., Goudou,D., Blicou,M., Robert,B. and  
Rieger,F.  
The HERV-W/7q family in the human genome. Potential for protein  
expression and gene regulation  
Cell. Mol. Biol. 48, 213-217 (2002)  
REFERENCE 3 (bases 1 to 1624)  
Goudou,D., Perin,J.P., Rieger,F., Robert,B. and Alliel,P.M.  
Direct Submission  
JOURNAL Submitted (29-APR-2002) U-488, INSERM, 80, rue du General Leclerc,  
Le Kremlin-Bicetre 94270, France  
FEATURES  
source  
1. 1624  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/tissue="placenta"  
25. >1624  
/codon\_start=1  
/product="envrin"  
/protein\_id="AA03413.1"  
/db\_xref="GI:20978308"  
/translation="MALPYHIFLFTVLLPSFTLLAPPCRCMTSSSPYOEFLMMPOR  
GNIDAPYSRSLSGKPTPTATHTHPRNCYHSATLCMHANTHYGKMINPSCPGGLV  
TVCWTVFTQGMDSGGVODARREKHVEVLSQTRVHGTSPPYKGLDLSKLTETLT  
HTRLVSLFNTTLGLHEVSAQNPNCMCICPLNRPVYSIPVPOMNNFSEINTTSV  
LVGLVSNLEITHRSNLTGVKFSNTYTTNSOCIRWVTPPOIICLPSGIFVCGTSA  
YRCLNGSSSMCEFLSPVPMPTIYEDDLVYVSKPRNRPVILPVIAGVIGALG  
TGIGITTSQFYKLSQELNGEMERVADSLVTLQDOLNSIAAVILNRRALDILTAE  
RGGTCFLGEECCYYNOSGIVTKVKEIRRIORRAEELNRPWGLLSQMMWILP  
FLGLAAILILLFLGPCIENLVNFVSSRIEAVKLQMPKQSKTKTKYRRPLDRPAS  
RSDVNDIKGTPPEIISAQPLRN"  
BASE COUNT 449 a 456 c 290 g 429 t  
ORIGIN

Query Match 100.0%; Score 201; DB 9; Length 1624;  
Best Local Similarity 100.0%; Freq. No. 1e-48;  
Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AACTTCAGCAGCAAGAAATAACACACCTCCGTTTAGTAGAGACCTCTGTTCCATCTG 60

|||||  
Db 646 AACTTACGACAGCAAAATAAACACACCTCCGTTTGTAGAGACCTCTGTTCCAAATCTG 705  
OY 61 GAAATAACCATACCTCAAACTCAGCTGTGTAAATTTGCAATACATACACAC 120  
Db 706 GAAATAACCATACCTCAAACTCAGCTGTGTAAATTTGCAATACATACACAC 765  
OY 121 AACTCCCAATGCATCAGGTGGGTAACTCCACACAAATAGTGTGCTACCTCAGGA 180  
Db 766 AACTCCCAATGCATCAGGTGGGTAACTCCACACAAATAGTGTGCTACCTCAGGA 825  
OY 181 AATATTTTGTCTGTGTAC 201  
Db 826 AATATTTTGTCTGTGTAC 846

RESULT 11  
AF513360 1860 bp mRNA linear PRI 05-JUN-2002  
LOCUS Homo sapiens enverlin mRNA, complete cds.  
DEFINITION AF513360  
ACCESSION AF513360  
VERSION AF513360.1 GI:21326140  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homindae; Homo.  
1 (bases 1 to 1860)  
Allel.P.M., Perlin,J.P., Plerig,R., Nussbaum,J.L., Menard,A. and  
Rieger,F.  
TITLE Endogenous retroviruses and multiple sclerosis. Part 2: HERV-7q and  
its env transcripts  
C. R. Acad. Sci. III, Sect. Vie 312, 857-863 (1998)  
2 (bases 1 to 1860)  
Allel.P.M., Perlin,J.P., Goudou,D., Bitoun,M., Robert,B. and  
Rieger,F.  
TITLE The HERV-7q family in the human genome. Potential for protein  
expression and gene regulation  
Cell Mol. Biol. 48 (2), 213-217 (2002)  
JOURNAL 21985840  
MEDLINE 11990458  
PUBMED  
REFERENCE 3 (bases 1 to 1860)  
Allel.P.M., Goudou,D., Perlin,J.P. and Rieger,F.  
TITLE Direct Submission  
Submitted (18-MAY-2002) U-488, INSERM, 80, rue du General Leclerc,  
Le Kremlin-Bicetre 94270, France  
FEATURES  
source  
1. 1860  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="7"  
/map="7q21"  
/tissue\_type="Placenta"  
59..1675  
/codon\_start=1  
/product="enverlin"  
/protein\_id="AA047599.1"  
/db\_xref="GI:21326141"  
/translation="MALPHTLFTVLPSFTLAPPPCRMTSSPYOELMRORP  
GNIDAPSYRLSKGPTPTAHMHPNRCNHSATLMAHATHTWTKMNPSCGSLG  
TVCMTYFTDTGMSDGGVODAREKHVKVILQLRVHSTSPYGLDLSKHEVLT  
HTRVLSLFTWTLGLHEVSAONPTNCWICLPLNFPYVSIIPVEOMNNSFEINTSY  
LVGPLVSNLEHTSNLNCVRFNSNTTYNSCIRRVPPPTOIVCLPSGIFVCGTSA  
YKLGSSSESMCELSFLVAPMTITTEODLYNVISKPRKRVPIIPVIGAGVIGAG  
TGIGITTSSTORYKYSOELNDMERVADSVITLDOCLAAVYLQNRALDLTAE  
RGCTCLFLGECCYVYNSGITYTEKVEKTRDIOKRAEELRTGFMGLSOWPMTLP  
FLGPLAAILLLFLGPCIFNLVNFVSSKIEAVKLOMEPKMOSKTKIYRPLDRPASP  
RSDVNDIKGTPEEISAQPLLRPSAGSS"

BASE COUNT 513 a 520 c 350 g 477 t  
ORIGIN

Query Match 100.0%; Score 201; DB 9; Length 1860;  
Best Local Similarity 100.0%; Pred. No. 1e-48;

Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 AACTTACGACAGCAAAATAAACACACCTCCGTTTGTAGAGACCTCTGTTCCAAATCTG 60  
Db 680 AACTTACGACAGCAAAATAAACACACCTCCGTTTGTAGAGACCTCTGTTCCAAATCTG 739  
OY 61 GAAATAACCATACCTCAAACTCAGCTGTGTAAATTTGCAATACATACACAC 120  
Db 740 GAAATAACCATACCTCAAACTCAGCTGTGTAAATTTGCAATACATACACAC 799  
OY 121 AACTCCCAATGCATCAGGTGGGTAACTCCACACAAATAGTGTGCTACCTCAGGA 180  
Db 800 AACTCCCAATGCATCAGGTGGGTAACTCCACACAAATAGTGTGCTACCTCAGGA 859  
OY 181 AATATTTTGTCTGTGTAC 201  
Db 860 AATATTTTGTCTGTGTAC 880

RESULT 12  
AX000960 1948 bp DNA linear PAT 10-MAR-2000  
LOCUS AX000960  
DEFINITION Sequence 5 from Patent WO9902656.  
ACCESSION AX000960  
VERSION AX000960.1 GI:7241202  
KEYWORDS  
SOURCE unidentified.  
ORGANISM unidentified.  
REFERENCE 1 (bases 1 to 1948)  
Beseme,F. and Blond,J.  
TITLE ENDOGENETIC RETROVIRAL SEQUENCES, ASSOCIATED WITH AUTOIMMUNE  
DISEASES OR WITH PREGNANCY DISORDERS  
JOURNAL Patent: WO 9902656-A 5 21-JAN-1999;  
BIO MERIEUX (FR); BESEME FREDERIC (FR)  
FEATURES  
source  
1. 1948  
/organism="unidentified"  
/db\_xref="taxon:32644"

BASE COUNT 505 a 549 c 381 g 510 t 3 others  
ORIGIN

Query Match 100.0%; Score 201; DB 6; Length 1948;  
Best Local Similarity 100.0%; Pred. No. 1e-48;  
Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 AACTTACGACAGCAAAATAAACACACCTCCGTTTGTAGAGACCTCTGTTCCAAATCTG 60  
Db 1340 AACTTACGACAGCAAAATAAACACACCTCCGTTTGTAGAGACCTCTGTTCCAAATCTG 1399  
OY 61 GAAATAACCATACCTCAAACTCAGCTGTGTAAATTTGCAATACATACACAC 120  
Db 1400 GAAATAACCATACCTCAAACTCAGCTGTGTAAATTTGCAATACATACACAC 1459  
OY 121 AACTCCCAATGCATCAGGTGGGTAACTCCACACAAATAGTGTGCTACCTCAGGA 180  
Db 1460 AACTCCCAATGCATCAGGTGGGTAACTCCACACAAATAGTGTGCTACCTCAGGA 1519  
OY 181 AATATTTTGTCTGTGTAC 201  
Db 1520 AATATTTTGTCTGTGTAC 1540

RESULT 13  
AX027474 1948 bp DNA linear PAT 16-SEP-2000  
LOCUS AX027474  
DEFINITION Sequence 24 from Patent FR2788784.  
ACCESSION AX027474  
VERSION AX027474.1 GI:10188438  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;



REFERENCE 1 (bases 1 to 1948)  
AUTHORS Mallet, F., Voisset, C. and Paranhos, B. G.  
JOURNAL Patent: FR 2788784-A 24 28-JUL-2000;  
BIO MERIEUX (FR)

FEATURES  
Source location/Qualifiers  
1..1948  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"

BASE COUNT 505 a 549 c 381 g 510 t 3 others

ORIGIN

Query Match 100.0%; Score 201; DB 6; Length 1948;  
Best Local Similarity 100.0%; Pred. No. 1e-48;  
Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AACTTCAGCAGAGAAATTAACACACCTCCGTTTATAGAGACCTCTGTTTCCAACTCG 60  
|||||  
DB 1340 AACTTCAGCAGAGAAATTAACACACCTCCGTTTATAGAGACCTCTGTTTCCAACTCG 1399  
|||||  
QY 61 GAATAAACCATACCTCAACCTGCTGTAATAATTGCAATACATACACCAACC 120  
|||||  
DB 1400 GAATAAACCATACCTCAACCTGCTGTAATAATTGCAATACATACACCAACC 1459  
|||||  
QY 121 AACTCCCAATGATCAGTGGGTAACTCTCCACACAAATAGTCTGCTACCTCAGGA 180  
|||||  
DB 1460 AACTCCCAATGATCAGTGGGTAACTCTCCACACAAATAGTCTGCTACCTCAGGA 1519  
|||||  
QY 181 ATATTTTGTCTGTGTGTTACC 201  
|||||  
DB 1520 ATATTTTGTCTGTGTGTTACC 1540  
|||||

RESULT 14  
AF072505 1948 bp mRNA linear PRI 10-FEB-1999  
LOCUS Homo sapiens endogenous retrovirus W envelope protein mRNA, partial  
DEFINITION cds.  
ACCESSION AF072505  
VERSION AF072505.1 GI:4262289  
KEYWORDS  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 1948)  
AUTHORS Blond, J. L., Beseme, F., Duret, L., Bouton, O., Bedin, F., Perron, H.,  
Mandrand, B. and Mallet, F.  
TITLE Molecular characterization and placental expression of HERV-W, a  
new human endogenous retrovirus family  
JOURNAL J. Virol. 73 (2), 1175-1185 (1999)  
PUBMED 9882319  
MEDLINE 99098005

REFERENCE 1 (bases 1 to 1948)  
AUTHORS Blond, J. L., Beseme, F. and Mallet, F.  
TITLE Direct Submission  
JOURNAL Submitted (19-JUN-1998) UM103 CNRS-Biomerieux, ENS Lyon, 46 allée  
d'Italie, Lyon, Cedex 07 69364, France

FEATURES  
Source location/Qualifiers  
1..1948  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_id="Cl.24.4"  
/tissue\_type="placenta"  
/clone\_lib="Clontech 5'-stretch plus library, Cat number  
HL5014a"  
1..1948  
/rpl\_family="HERV-W"  
<1..6  
/note="R"  
5'UTR 7..460  
/note="U5"  
primer\_bind 464..481

misc\_feature  
491..492  
/note="putative primer binding site for tRNA-W"

CDS  
719..>1948  
/codon\_start=1  
/product="envelope protein"  
/protein\_id="A01454.1"  
/db\_xref="GI:4262290"  
/translation="MALPYHIFLFTVSPSEFTLTAPPCCMTSSPYOEFWRMORP  
GNDAPSYRICKGPTPTFAHNTMPNRCYSAATLGMHANTHWTGKINDSPCGGIGV  
IVCWTFYTOTGMSDGGGVDDQAEKRYKVIISLTIVHGTSSPYKGLDLSKLEHETLT  
HRLVSLFNTLTGLHVSANPTNCWICLPLFRPVASIVPEQNNNESTELNTTIVS  
LVGPLVSNLEITHTSNLTCVKSNTTITNSQCIKRWVTPPTQIVCLPSGIFVCGTSA  
YRLINSSSESMCFSLFVPPMAIYTEODIYSVYSKPRNRYVILPEFVIGAGVLGALG  
TGIGITTSFOEYFKLSOELNGDMERVAISLVLDQQLNSLAIVLQNRRAIDLTLAE  
RGSTCLFLEECXYVNO"

BASE COUNT 505 a 549 c 381 g 510 t 3 others

ORIGIN

Query Match 100.0%; Score 201; DB 9; Length 1948;  
Best Local Similarity 100.0%; Pred. No. 1e-48;  
Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AACTTCAGCAGAGAAATTAACACACCTCCGTTTATAGAGACCTCTGTTTCCAACTCG 60  
|||||  
DB 1340 AACTTCAGCAGAGAAATTAACACACCTCCGTTTATAGAGACCTCTGTTTCCAACTCG 1399  
|||||  
QY 61 GAATAAACCATACCTCAACCTGCTGTAATAATTGCAATACATACACCAACC 120  
|||||  
DB 1400 GAATAAACCATACCTCAACCTGCTGTAATAATTGCAATACATACACCAACC 1459  
|||||  
QY 121 AACTCCCAATGATCAGTGGGTAACTCTCCACACAAATAGTCTGCTACCTCAGGA 180  
|||||  
DB 1460 AACTCCCAATGATCAGTGGGTAACTCTCCACACAAATAGTCTGCTACCTCAGGA 1519  
|||||  
QY 181 ATATTTTGTCTGTGTGTTACC 201  
|||||  
DB 1520 ATATTTTGTCTGTGTGTTACC 1540  
|||||

RESULT 15  
AX007999 2055 bp DNA linear PAT 06-SEP-2000  
LOCUS Sequence 22 from Patent WO967395.  
DEFINITION AX007999  
ACCESSION AX007999  
VERSION AX007999.1 GI:9995696  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 2055)  
AUTHORS Perin, J. P., Rieger, F. and Alliel, P. M.  
TITLE Nucleic sequence and deduced protein sequence family with human  
endogenous retroviral motifs, and their uses  
JOURNAL INST NAT SANTE RECH MED (FR); PERIN JEAN PIERRE (FR); RIEGER  
FRANCOIS (FR); ALLIEL PATRICK M (FR)  
PUBMED 9967395-A 22 29-DEC-1999;  
MEDLINE 9967395-A 22 29-DEC-1999;

FEATURES  
Source location/Qualifiers  
1..2055  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
<1..87  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAC07392.1"  
/db\_xref="GI:9995697"  
/translation="PKTANLVADITSXKQYVLTQLTGTPP"

BASE COUNT 576 a 574 c 376 g 529 t

ORIGIN

Query Match 100.0%; Score 201; DB 6; Length 2055;  
Best Local Similarity 100.0%; Pred. No. 1e-48;

Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AACTTCAGCAGCAAGATTAACACCACTTCGTTTGTAGAGACCTCTGTTCATCTG 60  
|||||  
Db 1012 AACTTCAGCAGCAAGATTAACACCACTTCGTTTGTAGAGACCTCTGTTCATCTG 1071  
|||||

QY 61 GAAATTAACCCATACCTCAACCTGCTGTAAATTTAGCAATATACATACACAC 120  
|||||  
Db 1072 GAAATTAACCCATACCTCAACCTGCTGTAAATTTAGCAATATACATACACAC 1131  
|||||

QY 121 AACTCCCATGCTCAGTGCGTAACTCTCCACACAAATAGTCTGCTACCTCAGGA 180  
|||||  
Db 1132 AACTCCCATGCTCAGTGCGTAACTCTCCACACAAATAGTCTGCTACCTCAGGA 1191  
|||||

QY 181 ATATTTTGTCTGTGTAC 201  
|||||  
Db 1192 ATATTTTGTCTGTGTAC 1212  
|||||

Search completed: May 2, 2003, 15:26:49  
Job time : 693 secs

GenCore version 5.1.5  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 2, 2003, 12:49:35 ; Search time 159.333 Seconds

(without alignments)  
2840.910 Million cell updates/sec

Title: US-09-719-554-3\_COPY\_8500\_8700

Perfect score: 201  
Sequence: 1 aactcagcagcagaataa.....tatttttctgtgttacc 201

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2185239 segs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

N\_Geneseq\_101002:\*

- 1: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1980.DAT:\*
- 2: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1981.DAT:\*
- 3: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1983.DAT:\*
- 4: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1984.DAT:\*
- 5: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1985.DAT:\*
- 6: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1986.DAT:\*
- 7: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1987.DAT:\*
- 8: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1988.DAT:\*
- 9: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1989.DAT:\*
- 10: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1990.DAT:\*
- 11: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1991.DAT:\*
- 12: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1992.DAT:\*
- 13: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1993.DAT:\*
- 14: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1994.DAT:\*
- 15: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1995.DAT:\*
- 16: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1996.DAT:\*
- 17: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1997.DAT:\*
- 18: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1998.DAT:\*
- 19: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA1999.DAT:\*
- 20: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA2000.DAT:\*
- 21: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA2001A.DAT:\*
- 22: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA2001B.DAT:\*
- 23: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA2002.DAT:\*
- 24: /SIDS2/gcgcdata/geneseq/geneseqn-emb1/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	201	100.0	1617	22	AAH20070
2	201	100.0	1948	20	AAH25659
3	201	100.0	1948	21	AAH59209
4	201	100.0	2055	21	ABN97948
5	201	100.0	2599	21	ABN97927
6	201	100.0	2781	22	AAH55630
7	201	100.0	2930	24	AAH24195
8	201	100.0	2946	20	AAH7526
9	201	100.0	2946	21	AAH59468

10	201	100.0	10499	21	ABN97929
11	201	100.0	56093	24	ABH61744
12	200.6	99.8	7582	21	AAH59215
13	199.4	99.2	2782	20	AAH25661
14	199.4	99.2	2782	21	AAH59211
15	199.4	99.2	2782	22	AAH20069
16	199.9	99.0	7582	20	AAH25665
17	193	96.0	1481	19	AAH43217
18	193	96.0	1481	20	AAH29703
19	193	96.0	1629	21	AAH96625
20	193	96.0	1799	21	ABN97931
21	193	96.0	2030	21	AAH63826
22	193	96.0	46340	21	ABN97978
23	191.4	95.2	902	23	AAH71722
24	191.4	95.2	1039	23	AAH84206
25	191.4	95.2	1341	23	AAH71719
26	191.4	95.2	1341	23	AAH84194
27	175.4	87.3	792	22	ABA50940
28	175.4	87.3	792	22	ABA68919
29	175.4	87.3	792	22	ABA35872
30	175.4	87.3	792	22	AAH17251
31	175.4	87.3	792	22	AAH43040
32	175.4	87.3	792	22	AAH23803
33	175.4	87.3	792	22	AAH49115
34	175.4	87.3	792	22	AAH09412
35	175.4	87.3	792	24	ABH17105
36	175.4	87.3	1894	22	ABA45822
37	175.4	87.3	1894	22	ABA56337
38	175.4	87.3	1894	22	ABA25978
39	175.4	87.3	1894	22	AAH04516
40	175.4	87.3	1894	22	AAH30018
41	175.4	87.3	1894	22	AAH14608
42	175.4	87.3	1894	22	AAH35960
43	175.4	87.3	1894	24	AAH04422
44	175.4	87.3	1894	24	ABH04589
45	175.4	87.3	2784	21	ABN97930

#### ALIGNMENTS

RESULT 1					
AAH20070	ID	AAH20070	standard; DNA: 1617 BP.		
XX	XX	AAH20070:			
AC	AC	08-AUG-2001	(first entry)		
DT	DT				
XX	XX				
DE	DE	HERV-W envelope protein G coding sequence.			
XX	XX	Human endogenous retrovirus; HERV-W; HERV; chromosome 7; env protein;			
KW	KW	envelope protein; multiple sclerosis-related superantigen; vaccine;			
KW	KW	surface antigen; transmembrane; multiple sclerosis; neuroprotective;			
KW	KW	antisense-therapy; autoimmune disorder; ds.			
XX	XX				
OS	OS	Human endogenous retrovirus.			
XX	XX				
FT	FT	Key	Location/Qualifiers		
FT	FT	CDS	1..1617		
FT	FT		/*tag= a		
FT	FT		/product= "HERV-W envelope protein G"		
XX	XX				
XX	XX	WO200131021-A1.			
XX	XX				
XX	XX	03-MAY-2001.			
XX	XX				
XX	XX	30-OCT-2000; 2000WO-EP10659.			
XX	XX				
XX	XX	28-OCT-1999; 99EP-0402690.			
XX	XX				
XX	XX	(UYGE-) UNIV GENEVE.			
XX	XX				

Human retroviral s  
Colon adenocarcino  
Human endogenous r  
Human endogenous r  
5' non coding, 3'  
HERV-W envelope pr  
Complete human end  
Multiple sclerosis  
Clone C15 from MSR  
DNA encoding an en  
Human retroviral s  
Nucleotide sequenc  
Human retroviral s  
DNA encoding novel  
DNA encoding novel  
DNA encoding novel  
DNA encoding novel  
Human breast cell  
Human foetal liver  
Probe #14338 for g  
Human brain expres  
Human bone marrow  
Probe #13736 for g  
Probe #17801 used  
Probe #9403 used t  
Human genome-deriv  
Human breast cell  
Human foetal liver  
Probe #4444 for ge  
Human brain expres  
Human bone marrow  
Probe #4541 for ge  
Probe #4666 used t  
Probe #4413 used t  
Human genome-deriv  
Human retroviral s

PI Conrad B, Mach B;  
XX WPI: 2001-316336/33.  
DR P-PSDB: AAB75138.  
XX  
PT New human retrovirus HERV-W ENV proteins/peptides having superantigen  
XX activity useful for diagnosing and treating multiple sclerosis -  
PS Claim 13; Fig 10; 94pp; English.  
XX  
CC On the basis of the PBS-t-RNA motif used for the classification of human  
CC endogenous retrovirus (HERVs) the full length endogenous provirus which  
CC was located on the long arm of human chromosome 7 (9q21-22) has been  
CC designated HERV-W. The present invention describes proteins or peptides  
CC (I) having superantigen (Sag) activity comprising the ENV protein (ENV)  
CC of HERV-W, the surface protein (SU) and transmembrane (TM) sub-units. (I)  
CC have neuroprotective activity, and can be used in: vaccines; antisense-  
CC therapy; and HERV-W Sag activity-inhibitors. (I) and encoding DNA/RNA are  
CC useful for diagnosing multiple sclerosis (MS) or HERV-W-associated  
CC disorders. (I) are also useful for identifying substances (and optionally  
CC recovering) capable of binding to a retroviral superantigen associated  
CC with MS, substances capable of blocking Sag activity and substances  
CC capable of blocking transcription or translation of HERV-W retroviral  
CC superantigen. A protein or peptide derived from (I), modified to be  
CC devoid of Sag activity and being capable of generating an immune response  
CC against HERV-W retroviral Sag is useful in therapy. Nucleic acid  
CC molecules encoding (I) are useful as vaccines against MS. Substances  
CC capable of blocking Sag activity, capable of binding to a retroviral  
CC superantigen associated with MS, or capable of blocking transcription or  
CC translation of HERV-W retroviral superantigen for use in treating or  
CC prevention of MS. (I) and nucleic acids encoding them are useful for  
CC diagnosing autoimmune disease. The present sequence encodes the  
CC specifically claimed envelope protein of HERV-W designated G.  
XX  
SQ Sequence 1617 BP: 442 A; 452 C; 296 G; 427 T; 0 other;  
Query Match 100.0%; Score 201; DB 22; Length 1617;  
Best Local Similarity 100.0%; Pred. No. 6.7e-52;  
Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 AACTTCAGCAGCAAAATAAACAACCACTCCGTTTGTAGAGACCTCTGTTCCAAATCG 60  
DB 622 AACTTCAGCAGCAAAATAAACAACCACTCCGTTTGTAGAGACCTCTGTTCCAAATCG 681  
OY 61 GAAATTAACCCATACCTCAAACTCAGCTGTGTAATTTAGCAATACTACATACACACACC 120  
DB 682 GAAATTAACCCATACCTCAAACTCAGCTGTGTAATTTAGCAATACTACATACACACACC 741  
OY 121 AACTCCCATGCATCAGTGGGTAACTCCTCCACACAAATAGTGCCTACCCCTAGGA 180  
DB 742 AACTCCCATGCATCAGTGGGTAACTCCTCCACACAAATAGTGCCTACCCCTAGGA 801  
OY 181 ATATTTTGTCTGTGTGAC 201  
DB 802 ATATTTTGTCTGTGTGAC 822  
RESULT 2  
AAK25659  
ID AAK25659 standard; CDNA to mRNA; 1948 BP.  
XX  
AC AAK25659;  
XX  
DT 21-MAY-1999 (first entry)  
XX  
DE Human endogenous retrovirus W clone cl.24.4.  
XX  
KM Clone; human endogenous retrovirus; genome; autoimmune disease;  
KM multiple sclerosis; rheumatoid polyarthritis; insulin-dependent diabetes;  
KM disseminated lupus erythematosus; pregnancy; chromosomal marker; ss.  
XX  
OS Human endogenous retrovirus.

XX  
PN WO9902696-A1.  
XX  
PD 21-JAN-1999.  
XX  
XX  
PF 06-JUL-1998; 98WO-FR01442.  
XX  
PR 07-JUL-1997; 97FR-0008815.  
XX  
PA (INMR) BIO MERIEUX.  
PI Beseme F, Blond JL, Bouton O, Mallet F, Mandrand B;  
XX WPI: 1999-120897/10.  
XX  
PT New nucleic acid sequences from human endogenous retrovirus-W -  
PT expressed exclusively in placenta and useful in diagnosis and  
PT therapy of autoimmune disease, and abnormal or failed pregnancy  
PS Claim 1; Page 56-58; 106pp; French.  
XX  
CC This sequence represents clone cl.24.4 of the human endogenous retrovirus  
CC (HERV) W genome. The nucleic acids, their fragments or peptides encoded  
CC by them are markers of autoimmune disease (e.g. multiple sclerosis,  
CC rheumatoid polyarthritis, disseminated lupus erythematosus, insulin-  
CC dependent diabetes and related pathologies) and of abnormal or  
CC unsuccessful pregnancy and can be used as chromosomal markers for  
CC susceptibility to these conditions, or proximity markers of genes  
CC associated with this susceptibility.  
XX  
SQ Sequence 1948 BP: 505 A; 549 C; 381 G; 510 T; 3 other;  
Query Match 100.0%; Score 201; DB 20; Length 1948;  
Best Local Similarity 100.0%; Pred. No. 7.1e-52;  
Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 AACTTCAGCAGCAAAATAAACAACCACTCCGTTTGTAGAGACCTCTGTTCCAAATCG 60  
DB 1340 AACTTCAGCAGCAAAATAAACAACCACTCCGTTTGTAGAGACCTCTGTTCCAAATCG 1399  
OY 61 GAAATTAACCCATACCTCAAACTCAGCTGTGTAATTTAGCAATACTACATACACACACC 120  
DB 1400 GAAATTAACCCATACCTCAAACTCAGCTGTGTAATTTAGCAATACTACATACACACACC 1459  
OY 121 AACTCCCATGCATCAGTGGGTAACTCCTCCACACAAATAGTGCCTACCCCTAGGA 180  
DB 1460 AACTCCCATGCATCAGTGGGTAACTCCTCCACACAAATAGTGCCTACCCCTAGGA 1519  
OY 181 ATATTTTGTCTGTGTGAC 201  
DB 1520 ATATTTTGTCTGTGTGAC 1540  
RESULT 3  
AAA59209  
ID AAA59209 standard; DNA; 1948 BP.  
XX  
AC AAA59209;  
XX  
DT 07-NOV-2000 (first entry)  
XX  
DE 5' non coding, 3' pol gene, and 5' env gene of HERV-W from human genome.  
XX  
KM Autoimmune disease; retrovirus; human endogenous retrovirus W; HERV-W;  
KM gag gene; pregnancy; multiple sclerosis; T cell proliferation; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200043521-A2.  
XX  
PD 27-JUL-2000.  
XX  
PF 21-JAN-2000; 2000WO-FR00144.

XX 21-JAN-1999; 99FR-0000888.  
PR  
XX  
XX (INRM ) BIO MERIEUX.  
PA  
XX  
XX Paranhos-Baccala G, Mallet F, Volisset C;  
PI  
XX WPI; 2000-499229/44.  
DR  
XX  
XX New nucleic acid from human endogenous retrovirus, useful e.g. for  
PT diagnosis of autoimmune disease and complications of pregnancy,  
PT contains at least part of the gag gene  
XX  
XX Disclosure; Page 45; 53pp; French.  
XX  
XX The present sequence represents an endogenous retroviral nucleic acid  
CC fragment, which is associated with an autoimmune disease, and is  
CC integrated into the human genome. The fragment is originally derived  
CC from a novel retrovirus, human endogenous retrovirus W (HERV-W). The  
CC HERV-W retrovirus is associated with autoimmune disease, failure of  
CC pregnancy or disorders of pregnancy. The nucleic acid fragment, or  
CC proteins derived from it, are useful for diagnosis of autoimmune  
CC disease (specifically multiple sclerosis) and for monitoring pregnancy.  
CC The nucleic acid fragments may also be used for in situ labelling of  
CC isolated chromosomes, while the transcription product can be used to  
CC study or monitor T cell proliferation in vitro.  
XX  
SQ Sequence 1948 BP; 505 A; 549 C; 381 G; 510 T; 3 other;  
Query Match 100.0%; Score 201; DB 21; Length 1948;  
Best Local Similarity 100.0%; Pred. No. 7.1e-52;  
Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 AACTTCAGCAGCAAAATTAACACCACTTCCTTTAGTAGACCTCTGTTCCAAATCTG 60  
DB 1340 AACTTCAGCAGCAAAATTAACACCACTTCCTTTAGTAGACCTCTGTTCCAAATCTG 1399  
OY 61 GAAATTAACCCATACCTCAAACTCAGCTGTGTAATAATTAGCAATACATACACAAACC 120  
DB 1400 GAAATTAACCCATACCTCAAACTCAGCTGTGTAATAATTAGCAATACATACACAAACC 1459  
OY 121 AACTCCCAATGATCAGGTGGGTAACTCTCCACACAAATAGTCTGCTACCTCAGAGA 180  
DB 1460 AACTCCCAATGATCAGGTGGGTAACTCTCCACACAAATAGTCTGCTACCTCAGAGA 1519  
OY 181 ATATTTTGTCTGTGTGATCC 201  
DB 1520 ATATTTTGTCTGTGTGATCC 1540

RESULT 4  
ABN97948  
ID ABN97948 standard; DNA; 2055 BP.  
XX  
XX AC ABN97948;  
XX  
XX DT 01-AUG-2002 (first entry)  
XX  
XX DE Human retroviral HERV-7q env coding sequence.  
XX  
XX KW Autoimmune disease; HERV-7q; chromosome 7q; immunotherapy;  
XX multiple sclerosis; ds.  
XX  
XX OS Human retrovirus.  
XX  
XX PN WO9967395-A1.  
XX  
XX PD 29-DEC-1999.  
XX  
XX PF 23-JUN-1999; 99WO-FR01513.  
XX  
XX PR 23-JUN-1998; 98FR-0007920.

PA (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.  
XX  
XX Alliel PM, Perin J, Rieger F;  
PI  
XX WPI; 2000-160587/14.  
DR  
XX  
XX New nucleic acid sequences of human endogenous retrovirus, HERV-7q,  
PT used for diagnosis, treatment and prevention of autoimmune and  
PT neurological diseases  
XX  
XX Claim 3; Page 142-145; 225pp; French.  
XX  
XX The present invention relates to new nucleic acid sequences of human  
CC endogenous retrovirus, HERV-7q, which is located on chromosome 7q.  
CC Regulatory elements associated with HERV-7q may alter expression of other  
CC genes (even remote genes) on the same chromosome, inducing immunological  
CC and/or neurological changes (which may be pathological or protective/  
CC curative). HERV-7q peptides can be used to improve efficiency of the  
CC immune response, e.g. in immunotherapy. HERV-7q peptides and their coding  
CC sequences can be used in immunogenic or vaccinating compositions, for  
CC protection against autoimmune diseases, particularly multiple sclerosis.  
CC The peptides may also be used (by sequence comparison) to detect/identify  
CC endogenous retroviruses that are abnormally expressed in cancer,  
CC neuropathologies or other autoimmune diseases. The present sequence was  
CC used to illustrate the invention.  
XX  
SQ Sequence 2055 BP; 576 A; 574 C; 376 G; 529 T; 0 other;  
Query Match 100.0%; Score 201; DB 21; Length 2055;  
Best Local Similarity 100.0%; Pred. No. 7.2e-52;  
Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 AACTTCAGCAGCAAAATTAACACCACTTCCTTTAGTAGACCTCTGTTCCAAATCTG 60  
DB 1012 AACTTCAGCAGCAAAATTAACACCACTTCCTTTAGTAGACCTCTGTTCCAAATCTG 1071  
OY 61 GAAATTAACCCATACCTCAAACTCAGCTGTGTAATAATTAGCAATACATACACAAACC 120  
DB 1072 GAAATTAACCCATACCTCAAACTCAGCTGTGTAATAATTAGCAATACATACACAAACC 1131  
OY 121 AACTCCCAATGATCAGGTGGGTAACTCTCCACACAAATAGTCTGCTACCTCAGAGA 180  
DB 1132 AACTCCCAATGATCAGGTGGGTAACTCTCCACACAAATAGTCTGCTACCTCAGAGA 1191  
OY 181 ATATTTTGTCTGTGTGATCC 201  
DB 1192 ATATTTTGTCTGTGTGATCC 1212

RESULT 5  
ABN97927  
ID ABN97927 standard; DNA; 2599 BP.  
XX  
XX AC ABN97927;  
XX  
XX DT 01-AUG-2002 (first entry)  
XX  
XX DE Human retroviral sequence HERV 7 env.  
XX  
XX KW Autoimmune disease; HERV-7q; chromosome 7q; immunotherapy;  
XX multiple sclerosis; ds.  
XX  
XX OS Human retrovirus.  
XX  
XX PN WO9967395-A1.  
XX  
XX PD 29-DEC-1999.  
XX  
XX PF 23-JUN-1999; 99WO-FR01513.  
XX  
XX PR 23-JUN-1998; 98FR-0007920.  
XX  
XX (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.

```
XX
PI Allel PM, Perin J, Rieger F;
XX
DR WPI: 2000-160587/14.
XX
PT New nucleic acid sequences of human endogenous retrovirus, HERV-7q,
PT used for diagnosis, treatment and prevention of autoimmune and
PT neurological diseases
XX
PS Claim 1: Page 128-129; 225pp; French.
XX
CC The present invention relates to new nucleic acid sequences of human
CC endogenous retrovirus, HERV-7q, which is located on chromosome 7q.
CC Regulatory elements associated with HERV-7q may alter expression of other
CC genes (even remote genes) on the same chromosome, inducing immunological
CC and/or neurological changes (which may be pathological or protective/
CC curative). HERV-7q peptides can be used to improve efficiency of the
CC immune response, e.g. in immunotherapy. HERV-7q peptides and their coding
CC sequences can be used in immunogenic or vaccinating compositions, for
CC protection against autoimmune diseases, particularly multiple sclerosis.
CC The peptides may also be used (by sequence comparison) to detect/identify
CC endogenous retroviruses that are abnormally expressed in cancer,
CC neuropathologies or other autoimmune diseases. The present sequence was
CC used to illustrate the invention.
XX
SQ Sequence 2599 BP; 744 A; 718 C; 495 G; 642 T; 0 other;
Query Match 100.0%; Score 201; DB 21; Length 2599;
Best Local Similarity 100.0%; Pred. No. 7.7e-52;
Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AACCTCAGCAGAGAAATAAACACCACTCCGTTTGTAGTAGGACCTCTGTTTCCAAATCG 60
DB 1550 AACCTCAGCAGAGAAATAAACACCACTCCGTTTGTAGTAGGACCTCTGTTTCCAAATCG 1609
QY 61 GAAATAGCCATACCTCAACCTCAACCTCGTAAATTTAGCAATACATACATACAGAAC 120
DB 1610 GAAATAGCCATACCTCAACCTCAACCTCGTAAATTTAGCAATACATACATACAGAAC 1669
QY 121 AACCTCCAAATGATCAGTGGTGAATCTCCACACAATAGTGTGCTACCCCTCAGGA 180
DB 1670 AACCTCCAAATGATCAGTGGTGAATCTCTCCACACAATAGTGTGCTACCCCTCAGGA 1729
QY 181 ATATTTTGTCTGTGTGATCC 201
DB 1730 ATATTTTGTCTGTGTGATCC 1750
RESULT 6
AAF55630
ID AAF55630 standard; DNA; 2781 BP.
XX
AC AAF55630;
XX
DT 29-MAY-2001 (first entry)
XX
DE Nucleotide sequence of a human endogenous retrovirus envelope protein.
XX
KW Envelope protein; HERV; syncytia formation; placental development;
XX syncytia; cancer; cell adhesion; ss.
XX
OS Human endogenous retrovirus.
XX
FH Key Location/Qualifiers
FT CDS 762..2378
FT /tag= a
FT /product= "envelope protein"
XX
PN WO200116171-A1.
XX
PD 08-MAR-2001.
XX
PF 01-SEP-2000; 2000MO-FR02429.
```

```
XX
PR 01-SEP-1999; 99FR-0011141.
PR 15-SEP-1999; 99FR-0011793.
XX
XX (INMR ) BIO MERIEUX.
PA (INMR ) INST NAT SANTE & RECH MEDICALE.
XX
PI Mallet F, Cosset F, Blond J, Lavillette D, Boulton O, Ruggieri A;
XX
DR WPI: 2001-226676/23.
DR P-PsDB: AAB67652.
XX
PT Detecting expression of human endogenous retrovirus envelope protein in
PT cells of a tissue or culture, from its ability to induce syncytia .
XX
PS Disclosure: Page 44-45; 57pp; French.
XX
CC The present sequence encodes a human endogenous retrovirus envelope
CC protein. The specification describes a method for detecting expression
CC of an envelope protein from a human endogenous retrovirus (HERV), in
CC cells, of a tissue or culture. The method comprises detecting syncytia
CC formation due to the fusogenic properties of the envelope protein.
CC Envelope polypeptides and polynucleotides are used to produce
CC therapeutic or prophylactic compositions, particularly for treatment of
CC cancer, to correct defects in placental development (or other natural
CC formation of other types of syncytia), and to promote adhesion of cells
CC in grafts or cellular repair processes. Expression of sequences
CC antisense to the polynucleotide are used to prevent formation of
CC syncytia.
XX
SQ Sequence 2781 BP; 747 A; 770 C; 556 G; 708 T; 0 other;
Query Match 100.0%; Score 201; DB 22; Length 2781;
Best Local Similarity 100.0%; Pred. No. 7.8e-52;
Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 AACCTCAGCAGAGAAATAAACACCACTCCGTTTGTAGTAGGACCTCTGTTTCCAAATCG 60
DB 1383 AACCTCAGCAGAGAAATAAACACCACTCCGTTTGTAGTAGGACCTCTGTTTCCAAATCG 1442
QY 61 GAAATAGCCATACCTCAACCTCAACCTCGTAAATTTAGCAATACATACATACAGAAC 120
DB 1443 GAAATAGCCATACCTCAACCTCAACCTCGTAAATTTAGCAATACATACATACAGAAC 1502
QY 121 AACCTCCAAATGATCAGTGGTGAATCTCCACACAATAGTGTGCTACCCCTCAGGA 180
DB 1503 AACCTCCAAATGATCAGTGGTGAATCTCTCCACACAATAGTGTGCTACCCCTCAGGA 1562
QY 181 ATATTTTGTCTGTGTGATCC 201
DB 1563 ATATTTTGTCTGTGTGATCC 1583
RESULT 7
AAD24195
ID AAD24195 standard; CDNA; 2930 BP.
XX
AC AAD24195;
XX
DT 07-MAY-2002 (first entry)
XX
DE Human syncytin cDNA.
XX
KW Human; syncytin; pre-eclampsia; gestational trophoblast disorder;
KW choriocarcinoma; hydatiform mole; placental site tumour; abortion;
KW envelope gene; human endogenous defective retrovirus; HERV-W, ss.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT CDS 930..2546
FT /tag= a
FT /product= "Syncytin"
```

XX MO200204678-A2.  
PN 17-JAN-2002.  
XX 09-JUL-2001; 2001WO-US21719.  
XX 07-JUL-2000; 2000US-216657P.  
XX (GEMV ) GENETICS INST INC.  
XX Keith JC, McCoy JM, Mi S;  
XX PI  
XX WPI: 2002-171727/22.  
XX P-PSDB; AAEL4540.  
XX Identifying a compound for treating a subject with or at risk of  
XX developing preeclampsia, comprises determining whether the expression  
XX of activity of syncytin in the cell is modulated in the presence of a  
XX test compound  
XX  
XX Disclosure: Page 39-42; 43pp; English.  
XX  
XX The invention relates to identifying compounds which are modulators  
XX of syncytin expression. The syncytin modulators are useful in diagnosis  
XX and treatment of preeclampsia and gestational trophoblast disorders (e.g.  
XX choriochorionoma, hydatiform mole, placental site tumour and missed/  
XX incomplete abortion). Syncytin is a human gene derived from the  
XX envelope gene of human endogenous defective retrovirus, HERV-W. The  
XX present invention is based partly on the discovery that syncytin  
XX expression is dramatically reduced in preeclampsia, and is also  
XX mis-localised to the apical syncytiotrophoblast membrane. The present  
XX sequence is human syncytin cDNA.  
XX  
SQ Sequence 2930 BP; 842 A; 800 C; 571 G; 717 T; 0 other;  
Query Match 100.0%; Score 201; DB 24; Length 2930;  
Best Local Similarity 100.0%; Pred. No. 7.9e-52;  
Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 AACTTCAGCAGAGAAATAAACACCACTTCGTTTATAGAGACCTCTGTTCCAAATCTG 60  
DB 1551 AACTTCAGCAGAGAAATAAACACCACTTCGTTTATAGAGACCTCTGTTCCAAATCTG 1610  
QY 61 GAATAATACCATACCTCAACCTCAGCTGTGTAAATTTAGCAATATATACACCAACC 120  
DB 1611 GAATAATACCATACCTCAACCTCAGCTGTGTAAATTTAGCAATATATACACCAACC 1670  
QY 121 AACTCCCAATGATCAGTGGGTAACTCTCCACACAAATAGTCTGCTACCCCTCAGGA 180  
DB 1671 AACTCCCAATGATCAGTGGGTAACTCTCCACACAAATAGTCTGCTACCCCTCAGGA 1730  
QY 181 ATATTTTGTGCTGTGCTGATCC 201  
DB 1731 ATATTTTGTGCTGTGCTGATCC 1751  
RESULT 8  
AA77526  
ID AAX77526 standard; cDNA; 2946 BP.  
XX AC AAX77526;  
XX 10-AUG-1999 (first entry)  
XX DE Human secreted protein AJ172\_2 cDNA.  
XX  
XX Secreted protein; testes; brain; blood; placenta; human; murine; thymus;  
XX bone marrow; treatment; prevention; nutrition; cytokine; immune; vaccine;  
XX cell proliferation; cell differentiation; suppressor; tumour inhibition;  
XX haematopoiesis regulator; activin; inhibin; chemotactic; chemokine;  
XX haemostatic; thrombolytic; receptor; ligand; anti-inflammatory; tumour;  
XX cadherin; tumour invasion suppressor; gene therapy; tissue growth; ss.

XX OS Homo sapiens.  
XX PN WO9926972-A1.  
XX 03-JUN-1999.  
XX 17-NOV-1998; 98WO-US24614.  
XX 20-OCT-1998; 98US-0175928.  
XX 21-NOV-1997; 97US-0976110.  
XX 18-MAY-1998; 98US-0080478.  
XX (GEMV ) GENETICS INST INC.  
XX Collins-Racie LA, Evans C, Jacobs K, Lavallie ER;  
XX McCoy JM, Merberg D, Treacy M;  
XX WPI: 1999-357813/30.  
XX P-PSDB; AAY08622.  
XX New polynucleotides encoding secreted proteins  
XX  
XX Claim 13a: Page 100-101; 142pp; English.  
XX  
XX This invention describes novel human secreted proteins encoded by  
XX polynucleotides isolated from human adult testes, adult brain, adult  
XX blood or adult placenta, or murine adult bone marrow or thymus cDNA  
XX libraries. The products of the invention are predicted to have biological  
XX activities which would make them suitable for treating, preventing or  
XX ameliorating medical conditions in humans and animals, although no  
XX supporting data is given. Suggested activities include nutritional  
XX activity, cytokine and cell proliferation/differentiation activity,  
XX immune stimulating (e.g. as vaccines) or suppressing activity,  
XX haematopoiesis regulating activity, tissue growth activity,  
XX activin/inhibin activity, chemotactic/chemokine activity, haemostatic  
XX and thrombolytic activity, receptor/ligand activity, anti-inflammatory  
XX activity, cadherin/tumour invasion suppressor activity, and tumour  
XX inhibition activity. The polynucleotides are also stated to be useful  
XX for gene therapy.  
XX  
SQ Sequence 2946 BP; 858 A; 801 C; 570 G; 717 T; 0 other;  
Query Match 100.0%; Score 201; DB 20; Length 2946;  
Best Local Similarity 100.0%; Pred. No. 8e-52;  
Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 AACTTCAGCAGAGAAATAAACACCACTTCGTTTATAGAGACCTCTGTTCCAAATCTG 60  
DB 1549 AACTTCAGCAGAGAAATAAACACCACTTCGTTTATAGAGACCTCTGTTCCAAATCTG 1608  
QY 61 GAATAATACCATACCTCAACCTCAGCTGTGTAAATTTAGCAATATATACACCAACC 120  
DB 1609 GAATAATACCATACCTCAACCTCAGCTGTGTAAATTTAGCAATATATACACCAACC 1668  
QY 121 AACTCCCAATGATCAGTGGGTAACTCTCCACACAAATAGTCTGCTACCCCTCAGGA 180  
DB 1669 AACTCCCAATGATCAGTGGGTAACTCTCCACACAAATAGTCTGCTACCCCTCAGGA 1728  
QY 181 ATATTTTGTGCTGTGCTGATCC 201  
DB 1729 ATATTTTGTGCTGTGCTGATCC 1749  
RESULT 9  
AA259468  
ID AA259468 standard; cDNA; 2946 BP.  
XX AC AA259468;  
XX 11-APR-2000 (first entry)  
XX DE Human secreted protein AJ172\_2 polynucleotide sequence.

```
XX Human; secreted protein; disease diagnosis; pre-eclampsia; cancer;
KW placental pathology; metastasis inhibition; nutritional activity;
KW immune stimulator; hematopoiesis regulator; tissue growth;
KW tumour inhibitor; anti-inflammatory; clone AJ172_2; ATCC_98115;
KW gene therapy; ss.
XX
XX Homo sapiens.
XX
XX WO960020-A1.
XX
XX 25-NOV-1999.
XX
XX 17-MAY-1999; 99WO-US10915.
XX
XX 18-MAY-1998; 98US-0080478.
XX
XX 20-OCT-1998; 98US-0175928.
XX
XX (GENE ) GENETICS INST INC.
XX
XX Jacobs K, McCoy JM, LaVallie ER, Collins-Racie LA, Evans C;
XX Merberg D, Ml S, Treacy M;
XX
XX WPI: 2000-116311/10.
XX
XX P-PSDB; AAI67313.
XX
XX
XX New polynucleotides encoding secreted cDNA libraries, used to develop
XX products for the diagnosis and treatment of neoplastic disease
XX
XX Claim 14; Page 107-108; 149pp; English.
XX
XX This is the human secreted protein AJ172_2 nucleotide sequence, obtained
XX from a human adult testes cDNA library. The invention relates to secreted
XX human and murine proteins. The polynucleotides and proteins are predicted
XX to have biological activities which would make them suitable for
XX treating, preventing or ameliorating medical conditions in humans and
XX animals. Detection of the levels of the proteins can be used for the
XX diagnosis of e.g. pre-eclampsia, placental pathology or cancer. Agents
XX which modulate the expression or function of the proteins may be used for
XX treating a neoplastic disease and inhibiting metastasis. Other suggested
XX activities include nutritional activity (e.g. in feeds), cytokine and
XX cell proliferation/differentiation activity, immune stimulating
XX (e.g. as vaccines) or suppressing activity, hematopoiesis regulating
XX activity, tissue growth activity, activin/inhibin activity,
XX chemotactic/chemokinetic activity, haemostatic and thrombolytic activity,
XX receptor/ligand activity, anti-inflammatory activity, cadherin/tumour
XX invasion suppressor activity, and tumour inhibition activity. The
XX polynucleotide sequences are also stated to be useful for gene therapy.
XX
XX Sequence 2946 BP; 858 A; 801 C; 570 G; 717 T; 0 other;
XX
XX Query Match 100.0%; Score 201; DB 21; Length 2946;
XX Best Local Similarity 100.0%; Pred. No. 8e-52;
XX Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 AACCTGACGACGAAATAAACACCACTCCGTTTAGTAGAGACCTCTGTTCCAAATCG 60
XX |||||||
XX DB 1549 AACCTGACGACGAAATAAACACCACTCCGTTTAGTAGAGACCTCTGTTCCAAATCG 1608
XX |||||||
XX 61 GAAATTAACCCATACCTCAACCTCAACCTGTTAAATTTAGCAATRACTACATACCAAC 120
XX |||||||
XX DB 1609 GAAATTAACCCATACCTCAACCTCAACCTGTTAAATTTAGCAATRACTACATACCAAC 1668
XX |||||||
XX 121 AACCTCCAAATGATGAGTGGTAACTCTCCACACAATATGCTGCTTACCCTCAGGA 180
XX |||||||
XX DB 1669 AACCTCCAAATGATGAGTGGTAACTCTCCACACAATATGCTGCTTACCCTCAGGA 1728
XX |||||||
XX 181 AATATTTTGTCTGTGTGATAC 201
XX |||||||
XX DB 1729 AATATTTTGTCTGTGTGATAC 1749
XX |||||||
XX RESULT 10
```

```
ABN97929
XX ID ABN97929 standard; DNA; 10499 BP.
XX
XX AC ABN97929;
XX
XX 01-AUG-2002 (first entry)
XX
XX Human retroviral sequence HERV-7q.
XX
XX Autoimmune disease; HERV-7q; chromosome 7q; immunotherapy;
XX multiple sclerosis; ds.
XX
XX Human retrovirus.
XX
XX WO967395-A1.
XX
XX 29-DEC-1999.
XX
XX 23-JUN-1999; 99WO-FR01513.
XX
XX 23-JUN-1998; 98FR-0007920.
XX
XX (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
XX
XX Alliel PM, Perin J, Rieger F;
XX
XX WPI: 2000-160587/14.
XX
XX New nucleic acid sequences of human endogenous retrovirus, HERV-7q,
XX used for diagnosis, treatment and prevention of autoimmune and
XX neurological diseases
XX
XX Claim 3; Fig 1; 225pp; French.
XX
XX The present invention relates to new nucleic acid sequences of human
XX endogenous retrovirus, HERV-7q, which is located on chromosome 7q.
XX Regulatory elements associated with HERV-7q may alter expression of other
XX genes (even remote genes) on the same chromosome, inducing immunological
XX and/or neurological changes (which may be pathological or protective/
XX curative). HERV-7q peptides can be used to improve efficiency of the
XX immune response, e.g. in immunotherapy. HERV-7q peptides and their coding
XX sequences can be used in immunogenic or vaccinating compositions, for
XX protection against autoimmune diseases, particularly multiple sclerosis.
XX The peptides may also be used (by sequence comparison) to detect/identify
XX endogenous retroviruses that are abnormally expressed in cancer,
XX neuropathologies or other autoimmune diseases. The present sequence was
XX used to illustrate the invention.
XX
XX Sequence 10499 BP; 3048 A; 2676 C; 2280 G; 2495 T; 0 other;
XX
XX Query Match 100.0%; Score 201; DB 21; Length 10499;
XX Best Local Similarity 100.0%; Pred. No. 1.1e-51;
XX Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 AACCTGACGACGAAATAAACACCACTCCGTTTAGTAGAGACCTCTGTTCCAAATCG 60
XX |||||||
XX DB 8500 AACCTGACGACGAAATAAACACCACTCCGTTTAGTAGAGACCTCTGTTCCAAATCG 8559
XX |||||||
XX 61 GAAATTAACCCATACCTCAACCTCAACCTGTTAAATTTAGCAATRACTACATACCAAC 120
XX |||||||
XX DB 8560 GAAATTAACCCATACCTCAACCTCAACCTGTTAAATTTAGCAATRACTACATACCAAC 8619
XX |||||||
XX 121 AACCTCCAAATGATGAGTGGTAACTCTCCACACAATATGCTGCTTACCCTCAGGA 180
XX |||||||
XX DB 8620 AACCTCCAAATGATGAGTGGTAACTCTCCACACAATATGCTGCTTACCCTCAGGA 8679
XX |||||||
XX 181 AATATTTTGTCTGTGTGATAC 201
XX |||||||
XX DB 8680 AATATTTTGTCTGTGTGATAC 8700
XX |||||||
XX RESULT 11
XX ABL61744
```



ID ABL61744 standard; DNA: 56093 BP.  
XX  
AC ABL61744;  
XX  
DT 15-MAY-2002 (first entry)  
XX  
DE Colon adenocarcinoma related gene sequence SEQ ID NO:81.  
XX  
KW Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;  
stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;  
cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;  
gene; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO200194629-A2.  
XX  
PD 13-DEC-2001.  
XX  
PF 30-MAY-2001; 2001WO-US10838.  
XX  
PR 05-JUN-2000; 2000US-209473P.  
PR 05-JUN-2000; 2000US-209531P.  
PR 18-SEP-2000; 2000US-233133P.  
PR 18-SEP-2000; 2000US-233617P.  
PR 20-SEP-2000; 2000US-234009P.  
PR 20-SEP-2000; 2000US-234034P.  
PR 20-SEP-2000; 2000US-234052P.  
PR 22-SEP-2000; 2000US-234509P.  
PR 22-SEP-2000; 2000US-234567P.  
PR 25-SEP-2000; 2000US-234923P.  
PR 25-SEP-2000; 2000US-234924P.  
PR 25-SEP-2000; 2000US-235077P.  
PR 25-SEP-2000; 2000US-235082P.  
PR 25-SEP-2000; 2000US-235134P.  
PR 25-SEP-2000; 2000US-235280P.  
PR 26-SEP-2000; 2000US-235637P.  
PR 26-SEP-2000; 2000US-235638P.  
PR 27-SEP-2000; 2000US-235711P.  
PR 27-SEP-2000; 2000US-235720P.  
PR 27-SEP-2000; 2000US-235840P.  
PR 27-SEP-2000; 2000US-235863P.  
PR 28-SEP-2000; 2000US-236028P.  
PR 28-SEP-2000; 2000US-236032P.  
PR 28-SEP-2000; 2000US-236033P.  
PR 28-SEP-2000; 2000US-236034P.  
PR 28-SEP-2000; 2000US-236109P.  
PR 28-SEP-2000; 2000US-236111P.  
PR 29-SEP-2000; 2000US-236842P.  
PR 29-SEP-2000; 2000US-236891P.  
PR 02-OCT-2000; 2000US-237172P.  
PR 02-OCT-2000; 2000US-237173P.  
PR 02-OCT-2000; 2000US-237278P.  
PR 02-OCT-2000; 2000US-237294P.  
PR 02-OCT-2000; 2000US-237295P.  
PR 02-OCT-2000; 2000US-237316P.  
PR 03-OCT-2000; 2000US-237425P.  
PR 03-OCT-2000; 2000US-237598P.  
PR 03-OCT-2000; 2000US-237604P.  
PR 03-OCT-2000; 2000US-237606P.  
PR 03-OCT-2000; 2000US-237608P.  
PR 01-NOV-2000; 2000US-244867P.  
PR 01-NOV-2000; 2000US-245084P.  
XX  
PA (AVAL-) AVALON PHARM.  
XX  
PI Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;  
PI Soppet DR, Weaver Z;  
XX  
XX WPI; 2002-188264/24.  
XX  
PT Screening for anti-neoplastic agent involves exposing cells to a  
PT chemical agent to be tested for anti-neoplastic activity, and

PT determining a change in expression of a gene of a signature gene set  
XX  
PS Claim 1; SEQ ID 81; 44pp; English.  
XX  
CC The present invention describes a method (M1) for screening for an  
CC anti-neoplastic agent. The method involves exposing cells to a chemical  
CC agent to be tested for anti-neoplastic activity, determining a change in  
CC expression of at least one gene (I) of a signature gene set, where (I)  
CC comprises a sequence (S) selected from 847 sequences (given in ABL61664  
CC to ABL70110), or is at least 95% identical to (S), where a change in  
CC expression is indicative of anti-neoplastic activity. (I) has cytostatic  
CC activity and can be used in gene therapy. M1 can be used for screening  
CC an anti-neoplastic agent, and can be used for producing a product which  
CC is the data collected with respect to the anti-neoplastic agent as a  
CC result of M1, and the data is sufficient to convey the chemical  
CC structure and/or properties of the agent. M1 can be used in the  
CC treatment of cancer such as colon, breast, stomach, lung, thyroid,  
CC oesophageal, ovarian, kidney, prostate or pancreatic cancer.  
CC adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,  
CC infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine  
CC carcinoma, papillary carcinoma and Wilm's tumour.  
XX  
SQ Sequence 56093 BP; 16164 A; 12346 C; 10702 G; 16881 T; 0 other;  
SQ  
Query Match 100.0%; Score 201; DB 24; Length 56093;  
Best Local Similarity 100.0%; Pred. No. 1.8e-51;  
Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 AACTTCAGCAGCAAGAAATTAACCCACCTCCCTTTAGTAGACCTCTGTTCCAACTCG 60  
|||||  
DB 36500 AACTTCAGCAGCAAGAAATTAACCCACCTCCCTTTAGTAGACCTCTGTTCCAACTCG 36559  
QY 61 GAATTAACCCATACCTCAACCTCACTGCTGTAATAATTAGCAATACATACCAACC 120  
|||||  
DB 36560 GAATTAACCCATACCTCAACCTCACTGCTGTAATAATTAGCAATACATACCAACC 36619  
QY 121 AACTCCCAATGATCAGTGGGTAGTCCCTCCACACAAATAGTGGCTGCTCCCAAGGA 180  
|||||  
DB 36620 AACTCCCAATGATCAGTGGGTAGTCCCTCCACACAAATAGTGGCTGCTCCCAAGGA 36679  
QY 181 ATATTTTTTGTCTGTGTAC 201  
|||||  
DB 36680 ATATTTTTTGTCTGTGTAC 36700  
RESULT 12  
AAA59215  
ID AAA59215 standard; DNA: 7582 BP.  
XX  
AC AAA59215;  
XX  
DT 07-NOV-2000 (first entry)  
XX  
DE Human endogenous retrovirus W (HERV-W) sequence.  
XX  
KW Autoimmune disease; retrovirus; human endogenous retrovirus W; HERV-W;  
gag gene; pregnancy; multiple sclerosis; T cell proliferation; ss.  
XX  
OS Human endogenous retrovirus.  
XX  
FH Key  
FT LTR 1..120 Location/Qualifiers  
FT FT /\*tag= a  
FT FT /note= "R of 5' LTR"  
FT LTR 121..575  
FT FT /\*tag= b  
FT FT /note= "U5 of 5' LTR"  
FT primer\_bind 579..596  
FT FT /\*tag= c  
FT CDS 581..7194  
FT FT /\*tag= d  
FT FT /note= "ORF1 env538"  
FT CDS 7039..7194

```

FT      /*tag- e
FT      /note- "ORF2 52 AA"
FT      7112..7255
FT      CDS      /*tag- f
FT      /note- "ORF3 48 AA"
FT      7244..7254
FT      misc_feature /*tag- g
FT      /note- "polypurine tract"
FT      7256..7582
FT      LTR      /*tag- h
FT      /note- "U3-R of 3' LTR"
FT      7563..7569
FT      polyA_signal /*tag- 1
FT
XX      WO200043521-A2.
XX      27-JUL-2000.
XX
XX      21-JAN-2000; 2000WO-FR00144.
XX      PR      21-JAN-1999; 99FR-0000888.
XX      PA      (INMR ) BIO MERIEUX.
XX      PI      Paranhos-Baccala G, Mallet F, Voisset C;
XX      DR      WPI; 2000-499229/44.
XX
XX      New nucleic acid from human endogenous retrovirus, useful e.g. for
XX      diagnosis of autoimmune disease and complications of pregnancy,
XX      PT      contains at least part of the gag gene
XX
XX      Disclosure: Page 49-52; 53pp; French.
XX
XX      The present sequence represents an endogenous retrovirus, which is
XX      CC      associated with an autoimmune disease, and is integrated into the human
XX      CC      genome. The retrovirus is human endogenous retrovirus W (HERV-W). The
XX      CC      HERV-W retrovirus is associated with autoimmune disease, failure of
XX      CC      pregnancy or disorders of pregnancy. HERV-W nucleic acid fragments, or
XX      CC      proteins derived from it, are useful for diagnosis of autoimmune
XX      CC      disease (specifically multiple sclerosis) and for monitoring pregnancy.
XX      CC      The nucleic acid fragments may also be used for in situ labelling of
XX      CC      isolated chromosomes, while the transcription product can be used to
XX      CC      study or monitor T cell proliferation in vitro.
XX
XX      Sequence 7582 BP; 2156 A; 1876 C; 1538 G; 1796 T; 216 other;
XX
XX      Query Match      99.8%; Score 200.6; DB 21; Length 7582;
XX      Best Local Similarity 99.5%; Pred. No. 1.4e-51;
XX      Matches 200; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

```

XX      21-MAY-1999 (first entry)
XX      DE      Human endogenous retrovirus W clone cl.FH74.
XX      KW      Clone; human endogenous retrovirus; genome; autoimmune disease;
XX      KW      multiple sclerosis; rheumatoid polyarthritis; insulin-dependent diabetes;
XX      KW      disseminated lupus erythematosus; pregnancy; chromosomal marker; ss.
XX      OS      Human endogenous retrovirus.
XX      PN      WO9902696-A1.
XX      PD      21-JAN-1999.
XX      PF      06-JUL-1998; 98WO-FR01442.
XX      PR      07-JUL-1997; 97FR-0008815.
XX      PA      (INMR ) BIO MERIEUX.
XX      PI      Beseme F, Blond JL, Bouton O, Mallet F, Mandrand B;
XX      DR      WPI; 1999-120897/10.
XX
XX      New nucleic acid sequences from human endogenous retrovirus-W -
XX      PT      expressed exclusively in placenta and useful in diagnosis and
XX      PT      therapy of autoimmune disease, and abnormal or failed pregnancy
XX      PS      Claim 1: Page 60-63; 106pp; French.
XX
XX      This sequence represents clone cl.PH74 of the human endogenous retrovirus
XX      CC      (HERV) W genome. The nucleic acids, their fragments or peptides encoded
XX      CC      by them are markers of autoimmune disease (e.g. multiple sclerosis,
XX      CC      rheumatoid polyarthritis, disseminated lupus erythematosus, insulin-
XX      CC      dependent diabetes and related pathologies) and of abnormal or
XX      CC      unsuccessful pregnancy and can be used as chromosomal markers for
XX      CC      susceptibility to these conditions, or proximity markers of genes
XX      CC      associated with this susceptibility.
XX
XX      Sequence 2782 BP; 741 A; 767 C; 565 G; 709 T; 0 other;
XX
XX      Query Match      99.2%; Score 199.4; DB 20; Length 2782;
XX      Best Local Similarity 99.5%; Pred. No. 2.4e-51;
XX      Matches 200; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

KW Autoimmune disease; retrovirus; human endogenous retrovirus W; HERV-W;  
XX gag gene; pregnancy; multiple sclerosis; T cell proliferation; ss.  
XX  
OS Homo sapiens.  
XX  
PN W0200043521-A2.  
XX  
PD 27-JUL-2000.  
XX  
PF 21-JAN-2000; 2000WO-FR00144.  
XX  
PR 21-JAN-1999; 99EP-0000888.  
XX  
PA (INMR ) BIO MERIEUX.  
XX  
PI Paranhos-Baccala G, Mallet F, Voisset C;  
XX  
DR WPI; 2000-499229/44.  
XX  
PT New nucleic acid from human endogenous retrovirus, useful e.g. for  
XX diagnosis of autoimmune disease and complications of pregnancy.  
PT contains at least part of the gag gene -  
XX  
PS Disclosure; Page 46-47; 53pp; French.  
XX  
CC The present sequence represents an endogenetic retroviral nucleic acid  
CC fragment, which is associated with an autoimmune disease, and is  
CC integrated into the human genome. The fragment is originally derived  
CC from a novel retrovirus, human endogenous retrovirus W (HERV-W). The  
CC HERV-W retrovirus is associated with autoimmune disease, failure of  
CC pregnancy or disorders of pregnancy. The nucleic acid fragment, or  
CC proteins derived from it, are useful for diagnosis of autoimmune  
CC disease (specifically multiple sclerosis) and for monitoring pregnancy.  
CC The nucleic acid fragments may also be used for in situ labelling of  
CC isolated chromosomes, while the transcription product can be used to  
CC study or monitor T cell proliferation in vitro.  
XX  
SQ Sequence 2782 BP; 741 A; 767 C; 565 G; 709 T; 0 other;

Query Match 99.28; Score 199.4; DB 21; Length 2782;  
Best Local Similarity 99.58; Pred. No. 2,4e-51;  
Matches 200; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1 AACTTCGACAGCAATTAACACCACTCCGTTAGTAGGAGCCTGTTGTTCCAACTG 60  
|||  
1384 AACTTCGACAGCAATTAACACCACTCCGTTAGTAGGAGCCTGTTGTTCCAACTG 1443

61 GAAATTAACCATACCTCAAACTCAGCTGTGTAAATTTAGCAATACATACACAACC 120  
|||  
1444 GAAATTAACCATACCTCAAACTCAGCTGTGTAAATTTAGCAATACATACACAACC 1503

121 AACCTCCAAATGACAGTGGGTAGTCCCTCCACACAATAGTCTGCTACCTCAGGA 180  
|||  
1504 AACCTCCAAATGACAGTGGGTAGTCCCTCCACACAATAGTCTGCTACCTCAGGA 1563

181 ATATTTTGTCTGTGTGTTAC 201  
|||  
1564 ATATTTTGTCTGTGTGTTAC 1584

RESULT 15  
AAH20069  
ID AAH20069 standard; DNA; 2782 BP.  
XX  
AC AAH20069;  
XX  
DT 08-AUG-2001 (first entry)  
XX  
DE HERV-W envelope protein G encoding nucleic acid.  
XX  
KW Human endogenous retrovirus; HERV-W; HERV; chromosome 7; env protein;  
KW envelope protein; multiple sclerosis-related superantigen; vaccine;  
KW surface antigen; transmembrane; multiple sclerosis; neuroprotective;

KW antisense-therapy; autoimmune disorder; ds.  
XX  
XX Human endogenous retrovirus.  
OS  
FH  
XX Key  
XX Location/Qualifiers  
XX 1..762  
XX /\*tag= a  
XX 5'UTR 763..2379  
XX CDS  
XX  
XX /\*tag= b  
XX /product= "HERV-W envelope protein G"  
XX /transl\_except= (pos:790..792,aa:Phe)  
XX /transl\_except= (pos:793..795,aa:Thr)  
XX /transl\_except= (pos:812..814,aa:Ileu)  
XX /transl\_except= (pos:818..820,aa:Ser)  
XX /transl\_except= (pos:865..867,aa:Tyr)  
XX /transl\_except= (pos:1174..1176,aa:Gln)  
XX /transl\_except= (pos:1441..1443,aa:Ileu)  
XX /transl\_except= (pos:1903..1905,aa:Gln)  
XX /transl\_except= (pos:2017..2019,aa:Iys)  
XX /transl\_except= (pos:2026..2028,aa:Arg)  
XX /transl\_except= (pos:2044..2046,aa:Arg)  
XX /transl\_except= (pos:2089..2091,aa:Gln)  
XX /transl\_except= (pos:2170..2172,aa:Asn)  
XX 2380..2782  
XX /\*tag= c  
XX  
XX W0200131021-A1.  
XX  
XX 03-MAY-2001.  
XX  
XX 30-OCT-2000; 2000WO-EP10659.  
XX  
XX 28-OCT-1999; 99EP-0402690.  
XX  
XX (UYGE-) UNIV GENEVE.  
XX  
XX Conrad B, Mach B;  
XX WPI; 2001-316336/33.  
XX P-PSDB: AAB/5138.  
XX  
XX New human retrovirus HERV-W ENV proteins/peptides having superantigen  
XX activity useful for diagnosing and treating multiple sclerosis -  
XX  
XX Claim 13; Fig 9; 94pp; English.  
XX  
XX On the basis of the PBS-t-RNA motif used for the classification of human  
XX endogenous retrovirus (HERVs) the full length endogenous provirus which  
XX was located on the long arm of human chromosome 7 (7q21-22) has been  
XX designated HERV-W. The present invention describes proteins or peptides  
XX (I) having superantigen (SAG) activity comprising the ENV protein (ENV)  
XX of HERV-W, the surface protein (SU) and transmembrane (TM) sub-units. (I)  
XX have neuroprotective activity, and can be used in: vaccines; antisense-  
XX therapy; and HERV-W SAG activity-inhibitors. (I) or HERV-W-associated  
XX useful for diagnosing multiple sclerosis (MS) or HERV-W-associated  
XX disorders. (I) are also useful for identifying substances (and optionally  
XX recovering) capable of binding to a retroviral superantigen associated  
XX with MS, substances capable of blocking SAG activity and substances  
XX capable of blocking transcription or translation of HERV-W retroviral  
XX superantigen. A protein or peptide derived from (I), modified to be  
XX devoid of SAG activity and being capable of generating an immune response  
XX against HERV-W retroviral SAG is useful in therapy. Nucleic acid  
XX molecules encoding (I) are useful as vaccines against MS. Substances  
XX capable of blocking SAG activity, capable of binding to a retroviral  
XX superantigen associated with MS, or capable of blocking transcription or  
XX translation of HERV-W retroviral superantigen for use in treating or  
XX preventing MS, obtained using (I) are useful for the treatment and  
XX prevention of MS. (I) and nucleic acids encoding them are useful for  
XX diagnosing autoimmune disease. The present sequence encodes the  
XX specifically claimed envelope protein of HERV-W designated G.  
XX  
SQ Sequence 2782 BP; 741 A; 768 C; 564 G; 709 T; 0 other;

```

Query Match          99.2%; Score 199.4; DB 22; Length 2782;
Best Local Similarity 99.5%; Pred. No. 2.4e-51;
Matches 200; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AACTTCAGCAGCAAAATAAACACCACTTCCGTTTGTAGTACGACCTCTGTGTTCCAAATCTG 60
    |||
Db 1384 AACTTCAGCAGCAAAATAAACACCACTTCCGTTTGTAGTACGACCTCTGTGTTCCAAATGTG 1443
    |||

QY 61 GAATAAACCCTACCTCAACCTCACCCTGTGTAATAATTAGCAATACATACACACACC 120
    |||
Db 1444 GAATAAACCCTACCTCAACCTCACCCTGTGTAATAATTAGCAATACATACACACACC 1503
    |||

QY 121 AACTCCCAATGATCAGTGGGTAACTCTCCACACAAATAGTGTGCTACCCCTCAGGA 180
    |||
Db 1504 AACTCCCAATGATCAGTGGGTAACTCTCCACACAAATAGTGTGCTACCCCTCAGGA 1563
    |||

QY 181 ATATTTTGTCTGTGTACC 201
    |||
Db 1564 ATATTTTGTCTGTGTACC 1584
    |||

```

Search completed: May 2, 2003, 14:51:44  
 Job time : 184.333 secs

GenCore version 5.1.5  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 2, 2003, 14:19:30 ; Search time 1112 Seconds  
(without alignments)  
2927.422 Million cell updates/sec

Title: US-09-719-554-3\_COPY\_8500\_8700

Perfect score: 201

Sequence: 1 aactcagcagcagaataaa.....tattttgtctgtgtacc 201

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*  
1: em\_estba:\*  
2: em\_esthum:\*  
3: em\_estin:\*  
4: em\_estmu:\*  
5: em\_estov:\*  
6: em\_estpl:\*  
7: em\_estro:\*  
8: em\_hcc:\*  
9: gb\_est1:\*  
10: gb\_est2:\*  
11: gb\_hcc:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vrt:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_trod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	201	100.0	459	14	R68740
2	201	100.0	723	9	AU138405
3	193	96.0	631	17	AQ261133
4	190	94.5	702	9	AU138097
5	171.2	85.2	346	13	BI053386
6	171	85.1	315	13	BI055459

C	7	171	85.1	315	13	BI056080	BI056080 RC5-GN028
C	8	171	85.1	376	13	BI056595	BI056595 RC5-GN028
C	9	170.8	85.0	374	14	T93661	T93661 Y605809.r1
C	10	169.4	84.3	345	13	BI056592	BI056592 RC5-GN028
C	11	169	84.1	313	13	BI056596	BI056596 RC5-GN028
C	12	169	84.1	373	13	BI052559	BI052559 RC5-GN028
C	13	169	84.1	373	13	BI052567	BI052567 RC5-GN028
C	14	169	84.1	379	13	BI052566	BI052566 RC5-GN028
C	15	169	84.1	382	13	BI052565	BI052565 RC5-GN028
C	16	167.4	83.3	373	13	BI05589	BI05589 RC5-GN028
C	17	167.4	83.3	374	13	BI056593	BI056593 RC5-GN028
C	18	165.8	82.5	300	13	BI055456	BI055456 RC5-GN028
C	19	165.8	82.5	300	13	BI056077	BI056077 RC5-GN028
C	20	165.8	82.5	373	13	BI052561	BI052561 RC5-GN028
C	21	165.8	82.5	373	13	BI052571	BI052571 RC5-GN028
C	22	165.8	82.5	373	13	BI052572	BI052572 RC5-GN028
C	23	164.2	81.7	373	13	BI052570	BI052570 RC5-GN028
C	24	158.8	79.0	331	13	BI055458	BI055458 RC5-GN028
C	25	158.8	79.0	331	13	BI056079	BI056079 RC5-GN028
C	26	154.8	77.0	334	13	BI053385	BI053385 RC5-GN028
C	27	116.8	58.1	586	10	BE019603	BE019603 ba84f03.y
C	28	107	53.2	474	14	R00743	R00743 Y67409.r1
C	29	101	50.2	267	14	H13097	H13097 Y110508.r1
C	30	100	49.8	714	12	BE734284	BE734284 601565487
C	31	91	45.3	128	14	T39519	T39519 Y807504.r1
C	32	84	41.8	586	17	A0776289	A0776289 HS_5528.A
C	33	81	40.3	790	12	BS572445	BS572445 602593450
C	34	79.6	39.6	718	9	AL601847	AL601847 DKE2P313K
C	35	69.4	34.5	317	13	BI056590	BI056590 RC5-GN028
C	36	39	19.4	349	14	BQ366047	BQ366047 RC6-GN007
C	37	37.4	18.6	939	17	CNS00CNG	AL059400 Drosophila
C	38	36.6	18.2	717	13	BM638420	BM638420 170006875
C	39	36.2	18.0	883	17	AZ671602	AZ671602 ENTIC17TF
C	40	36.2	18.0	919	17	AQ749036	AQ749036 HS_5575.A
C	41	36.2	18.0	922	17	A2686568	A2686568 ENTKQ46TR
C	42	35.8	17.8	732	17	CNS0113T	AL099769 Drosophila
C	43	35.6	17.7	430	9	AA058696	AA058696 zf58B04.r
C	44	35.6	17.7	1101	17	CNS00LT2	AL078714 Drosophila
C	45	17.4	640	10	10	BB538097	BB538097 BB538097

#### ALIGNMENTS

RESULT 1  
R68740  
LOCUS  
DEFINITION  
IMAGE:139258 5' similar to NP:B44282 B44282 RETROVIRUS-RELATED ENV  
POLYPROTEIN - ; mRNA sequence.  
R68740 459 bp mRNA linear EST 26-JUN-1995  
Y114906.r1 Soares placenta NP:B44282 B44282 RETROVIRUS-RELATED ENV  
POLYPROTEIN - ; mRNA sequence.

ACCESSION  
R68740  
VERSION  
R68740.1 GI:842257  
KEYWORDS  
EST.

SOURCE  
ORGANISM  
human.

REFERENCE  
AUTHORS

Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevaskis, E., Waterston, R., Williamson, A., Wobidmann, P. and Wilson, R.  
The WashU-Merck EST Project  
Unpublished (1995)  
Contact: Wilson RK  
Washington University School of Medicine  
444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@wustl.edu  
Insert Size: 1644  
High quality sequence stops: 316  
Source: IMAGE Consortium, LLNL

TITLE  
JOURNAL  
COMMENT

HRI human cDNA project  
Unpublished (2000)  
Contact: Takao Isogai  
Genomics Laboratory  
Helix Research Institute  
1512-3 Yana, Kasasazizu, Chiba 292-0812, Japan  
Tel: 81-438-52-3975  
Fax: 81-438-52-3986  
Email: genomics@hri.co.jp

Department of Eukaryotic Genomics  
The Institute for Genomic Research  
9712 Medical Center Dr., Rockville, MD 20850, USA  
Tel: 301 838 0200  
Fax: 301 838 0208  
Email: mdadams@tigr.org  
Clones are available from Research Genetics (info@resgen.com). BAC  
end search page: [http://www.tigr.org/tdb/humgen/bac\\_end\\_search/bac\\_end\\_search.html](http://www.tigr.org/tdb/humgen/bac_end_search/bac_end_search.html).  
Seq primer: M13-21  
Class: BAC ends.

```
/organism="Homo sapiens"  
/db_xref="taxon:9606"  
/clone="2506C15"
```

```
/clone_id="CITR-EI"
/sex="male"
/cell_type="sperm"
/note="Vector: pBelobAC11; Site_1: EcoRI; Site_2: EcoRI;
Caltech Human BAC library D"
BASE COUNT      182 a      189 c      99 g      161 t
ORIGIN

Query Match      96.0%; Score 193; DB 17; Length 631;
Best Local Similarity 97.5%; Pred. No. 4.6e-50;
Matches 196; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AACTTCAGCAGAAATAAACACACCTCCCTTTAGTAGAGACCTCTGTTCCATCTG 60
    |||||||
Db 245 AACTTCAGCAGAAATAAACACACCTCCCTTTAGTAGAGACCTCTGTTCCATCTG 304
    |||||||

QY 61 GAAATTAACCCATACCTCAACCTGCTGTAAATTTAGCAATACATACACACAC 120
    |||||||
Db 305 GAAATTAACCCATACCTCAACCTGCTGTAAATTTAGCAATACATACACACAC 364
    |||||||

QY 121 AACGCCAATGCATGAGTGGTACCTCCACACAAATAGTCTGCTACCTCAGAGA 180
    |||||||
Db 365 AGCTCCCAATGCATGAGTGGTACCTCCACACAAATAGTCTGCTACCTCAGAGA 424
    |||||||

QY 181 ATATTTTGTCTGTGTAC 201
    |||||||
Db 425 ATATTTTGTCTGTGTAC 445
    |||||||

RESULT 4
AUI38097      702 bp      mRNA      linear      EST 02-AUG-2002
LOCUS        AUI38097 PLACE1 Homo sapiens cDNA clone PLACE1007839 5', mRNA
DEFINITION   sequence.
ACCESSION    AUI38097
VERSION      AUI38097.1 GI:10999618
KEYWORDS     EST.
SOURCE       human.
ORGANISM     Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
              1 (bases 1 to 702)
REFERENCE     Ota,T., Nishikawa,T., Suzuki,Y., Ishii,S., Saito,K., Kawai,Y.,
              Yamamoto,J., Wakamatsu,A., Nakamura,Y., Nagai,T., Sugano,S. and
              Isogai,T.
              HRI human cDNA project
              Unpublished (2000)
              Contact: Takao Isogai
              Genomics Laboratory
              Helix Research Institute
              1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
              Tel: 81-438-52-3975
              Fax: 81-438-52-3986
              Email: genomics@hri.co.jp
              HRI human cDNA project; 5'- & 3'-end one pass sequencing; Helix
              Research Institute; cDNA library construction; Department of
              Virology, Institute of Medical Science, University of Tokyo, and
              Helix Research Institute.
              Location/Qualifiers
                source
                  1..702
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /clone_id="PLACE1007839"
                    /clone_id="PLACE1"
                    /tissue_type="placenta"
                    /note="Vector: pME18SF13"
BASE COUNT      206 a      173 c      131 g      189 t      3 others
ORIGIN

Query Match      94.5%; Score 190; DB 9; Length 702;
Best Local Similarity 100.0%; Pred. No. 4.2e-49;
Matches 190; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 12 AGAATATAACACCACTCCCTTTAGTAGAGACCTCTGTTCCATCTGAAATACCA 71
    |||||||
Db 1 AGAATATAACACCACTCCCTTTAGTAGAGACCTCTGTTCCATCTGAAATACCA 60
    |||||||

QY 72 TACCTCAACCTCACTCTGTAAATTTAGCAATACATACACACCACTCCCATG 131
    |||||||
Db 61 TACCTCAACCTCACTCTGTAAATTTAGCAATACATACACACCACTCCCATG 120
    |||||||

QY 132 CATGAGTGGGTAACTCTCCACACAATATGTCTGCTACCTCAGAGAAATTTTGT 191
    |||||||
Db 121 CATGAGTGGGTAACTCTCCACACAATATGTCTGCTACCTCAGAGAAATTTTGT 180
    |||||||

QY 192 CTGTGTGTAC 201
    |||||||
Db 181 CTGTGTGTAC 190
    |||||||

RESULT 5
BI053386/c      346 bp      mRNA      linear      EST 15-JUN-2001
LOCUS        BI053386/c
DEFINITION   RC5-GN0282-250101-011-E07 GN0282 Homo sapiens cDNA, mRNA sequence.
ACCESSION    BI053386
VERSION      BI053386.1 GI:14460916
KEYWORDS     EST.
SOURCE       human.
ORGANISM     Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
              1 (bases 1 to 346)
REFERENCE     Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
              Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
              Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
              Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
              ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
              Simpson,A.J.
              Shotgun sequencing of the human transcriptome with ORF expressed
              sequence tags
              Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
              2020263
              Contact: Simpson A.J.G.
              Laboratory of Cancer Genetics
              Ludwig Institute for Cancer Research
              Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
              Brazil
              Tel: +55-11-2704922
              Fax: +55-11-2707001
              Email: asimpson@ludwig.org.br
              This sequence was derived from the FAPESP/LICR Human Cancer Genome
              project. This entry can be seen in the following URL
              (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=RC5&st=2-RC5-GN0282-
              250101-011-E07&t3=2001-01-25&t4=1)
              Seq primer: puc 18 forward
              High quality sequence stop: 342.
              Location/Qualifiers
                source
                  1..346
                    /organism="Homo sapiens"
                    /db_xref="taxon:9606"
                    /clone_id="GN0282"
                    /dev_stage="Adult"
                    /note="Organ: placenta.normal; Vector: puc18; Site_1: SmaI
                    ; Site_2: SmaI; A mini-library was made by cloning
                    products derived from ORESTES PCR (U.S. Letters Patent
                    application No. 196,716 - Ludwig Institute for Cancer
                    Research) profiles into the pUC 18 vector. Reverse
                    transcription of tissue mRNA and cDNA amplification were
                    performed under low stringency conditions."
BASE COUNT      84 a      48 c      116 g      98 t
ORIGIN

Query Match      85.2%; Score 171.2; DB 13; Length 346;
Best Local Similarity 98.3%; Pred. No. 2.9e-43;
Matches 173; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
```

QY 1 AACTTCAGCAGAAATAAACACCACTTCCTTTAGTAGACCTCTGTTTCCANTGTG 60  
|||||  
Db 187 AACTTCAGCAGAAATAAACACCACTTCCTTTAGTAGACCTCTGTTTCCANTGTG 128  
QY 61 GAATTAACCCATACCTCAAACTCAGTGTAAATTTAGCAATACATACATACCAACC 120  
|||||  
Db 127 GAATTAACCCATACCTCAAACTCAGTGTAAATTTAGCAATACATACATACCAACC 68  
QY 121 AACTCCCAATGATCAGGTGGTAACTCTCCACACAAATATGCTGCTACCTTC 176  
|||||  
Db 67 AACTCCCAATGATCAGGTGGTAACTCTCCACACAAATATGCTGCTACCTTC 12

RESULT 6  
BI055459/c 315 bp mRNA linear EST 15-JUN-2001  
LOCUS BI055459  
DEFINITION RC5-GN0282-060201-012-H05 GN0282 Homo sapiens cDNA, mRNA sequence.  
ACCESSION BI055459  
VERSION BI055459.1 GI:14462989  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 315)  
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,  
Nagal,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,  
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,  
Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare  
M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and  
Simpson,A.J.  
TITLE Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
MEDLINE 20202663  
COMMENT Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?l=RC5&t2=RC5-GN0282-  
060201-012-H05&t3=2001-02-06&t4=1)  
Seq primer: puc 18 forward  
High quality sequence start: 11  
High quality sequence stop: 242.  
Location/Qualifiers  
FEATURES  
source  
1..315  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="GN0282"  
/dev\_stage="Adult"  
/note="Organ: placenta.normal; Vector: puc18; Site\_1: SmaI  
; Site\_2: SmaI; A mini-library was made by cloning  
products derived from ORESTES PCR (U.S. Letters Patent  
Application No. 196,716 - Ludwig Institute for Cancer  
Research) profiles into the puc 18 vector. Reverse  
transcription of tissue mRNA and cDNA amplification were  
performed under low stringency conditions."

BASE COUNT 74 a 49 c 95 g 95 t 2 others

ORIGIN  
Query Match 85.1%; Score 171; DB 13; Length 315;  
Best local Similarity 100.0%; Pred. No. 3.3e-43;  
Matches 171; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AACTTCAGCAGAAATAAACACCACTTCCTTTAGTAGACCTCTGTTTCCANTGTG 60  
|||||  
Db 179 AACTTCAGCAGAAATAAACACCACTTCCTTTAGTAGACCTCTGTTTCCANTGTG 120

QY 61 GAATTAACCCATACCTCAAACTCAGTGTAAATTTAGCAATACATACATACCAACC 120  
|||||  
Db 119 GAATTAACCCATACCTCAAACTCAGTGTAAATTTAGCAATACATACATACCAACC 60  
QY 121 AACTCCCAATGATCAGGTGGTAACTCTCCACACAAATATGCTGCTTA 171  
|||||  
Db 59 AACTCCCAATGATCAGGTGGTAACTCTCCACACAAATATGCTGCTTA 9

RESULT 7  
BI056080/c 315 bp mRNA linear EST 15-JUN-2001  
LOCUS BI056080  
DEFINITION RC5-GN0282-150201-012-H05 GN0282 Homo sapiens cDNA, mRNA sequence.  
ACCESSION BI056080  
VERSION BI056080.1 GI:14463610  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 315)  
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,  
Nagal,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,  
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,  
Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare  
M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and  
Simpson,A.J.  
TITLE Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
MEDLINE 20202663  
COMMENT Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?l=RC5&t2=RC5-GN0282-  
150201-012-H05&t3=2001-02-15&t4=1)  
Seq primer: puc 18 forward  
High quality sequence start: 11  
High quality sequence stop: 242.  
Location/Qualifiers  
FEATURES  
source  
1..315  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="GN0282"  
/dev\_stage="Adult"  
/note="Organ: placenta.normal; Vector: puc18; Site\_1: SmaI  
; Site\_2: SmaI; A mini-library was made by cloning  
products derived from ORESTES PCR (U.S. Letters Patent  
Application No. 196,716 - Ludwig Institute for Cancer  
Research) profiles into the puc 18 vector. Reverse  
transcription of tissue mRNA and cDNA amplification were  
performed under low stringency conditions."

BASE COUNT 74 a 49 c 95 g 95 t 2 others

ORIGIN  
Query Match 85.1%; Score 171; DB 13; Length 315;  
Best local Similarity 100.0%; Pred. No. 3.3e-43;  
Matches 171; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AACTTCAGCAGAAATAAACACCACTTCCTTTAGTAGACCTCTGTTTCCANTGTG 60  
|||||  
Db 179 AACTTCAGCAGAAATAAACACCACTTCCTTTAGTAGACCTCTGTTTCCANTGTG 120

QY 61 GAATTAACCCATACCTCAAACTCAGTGTAAATTTAGCAATACATACATACCAACC 120  
|||||





QY 179 GAATATTTTGTCTGTG 197  
|||||  
Db 298 GAATATTTTGTCTGTG 316  
|||||

RESULT 10  
BI056592/c 345 bp mRNA linear EST 15-JUN-2001  
LOCUS RC5-GN0282-210201-013-E03 GN0282 Homo sapiens cDNA, mRNA sequence.  
DEFINITION BI056592  
ACCESSION BI056592  
VERSION BI056592.1 GI:14464122  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 345)  
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,  
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,  
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H.,  
Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare  
, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and  
Simpson, A.J.  
Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags  
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

TITLE  
JOURNAL  
MEDLINE  
COMMENT  
Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?l=RC5&t2=RC5-GN0282-  
210201-013-E03&t3=2001-02-21&t4=1)  
Seq primer: puc 18 forward  
High quality sequence stop: 344.  
Location/Qualifiers  
1. 345  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="GN0282"  
/dev\_stage="Adult"  
/note="Organ: placenta,normal; Vector: puc18; Site\_1: SmaI  
; Site\_2: SmaI; A mini-library was made by cloning  
products derived from ORESTES PCR (U.S. Letters Patent  
application No. 196/716 - Ludwig Institute for Cancer  
Research) profiles into the puc 18 vector. Reverse  
transcription of tissue mRNA and cDNA amplification were  
performed under low stringency conditions."

BASE COUNT 85 a 45 c 107 g 108 t  
ORIGIN

Query Match 84.3%; Score 169.4; DB 13; Length 345;  
Best Local Similarity 99.4%; Pred. NO. 1.1e-42;  
Matches 170; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 AACTTCAGCAGAAATAAACACACTTCGTTTGTAGTAGACCTCTTGTTCATCTG 60  
|||||  
Db 172 AACTTCAGCAGAAATAAACACACTTCGTTTGTAGTAGACCTCTTGTTCATCTG 113  
|||||

QY 61 GAAATTAACCATCTCAAACTCACTGTTGTAATTTAGCAATATCTATACACAC 120  
|||||  
Db 112 GAAATTAACCATCTCAAACTCACTGTTGTAATTTAGCAATATCTATACACAC 53  
|||||

QY 121 AACTCCCATGATCAGTGAGTGAATCTCTCCACACAATATGTCTGCTTA 171  
|||||  
Db 52 AACTCCCATGATCAGTGAGTGAATCTCTCCACACAATATGTCTGCTTA 2  
|||||

RESULT 11  
BI056596/c 313 bp mRNA linear EST 15-JUN-2001  
LOCUS RC5-GN0282-210201-013-H02 GN0282 Homo sapiens cDNA, mRNA sequence.  
DEFINITION BI056596  
ACCESSION BI056596  
VERSION BI056596.1 GI:14464126  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 313)  
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,  
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,  
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H.,  
Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare  
, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and  
Simpson, A.J.  
Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags  
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

TITLE  
JOURNAL  
MEDLINE  
COMMENT  
Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?l=RC5&t2=RC5-GN0282-  
210201-013-H02&t3=2001-02-21&t4=1)  
Seq primer: puc 18 forward  
High quality sequence stop: 312.  
Location/Qualifiers  
1. 313  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="GN0282"  
/dev\_stage="Adult"  
/note="Organ: placenta,normal; Vector: puc18; Site\_1: SmaI  
; Site\_2: SmaI; A mini-library was made by cloning  
products derived from ORESTES PCR (U.S. Letters Patent  
application No. 196/716 - Ludwig Institute for Cancer  
Research) profiles into the puc 18 vector. Reverse  
transcription of tissue mRNA and cDNA amplification were  
performed under low stringency conditions."

BASE COUNT 77 a 46 c 99 g 91 t  
ORIGIN

Query Match 84.1%; Score 169; DB 13; Length 313;  
Best Local Similarity 100.0%; Pred. NO. 1.4e-42;  
Matches 169; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AACTTCAGCAGAAATAAACACACTTCGTTTGTAGTAGACCTCTTGTTCATCTG 60  
|||||  
Db 169 AACTTCAGCAGAAATAAACACACTTCGTTTGTAGTAGACCTCTTGTTCATCTG 110  
|||||

QY 61 GAAATTAACCATCTCAAACTCACTGTTGTAATTTAGCAATATCTATACACAC 120  
|||||  
Db 109 GAAATTAACCATCTCAAACTCACTGTTGTAATTTAGCAATATCTATACACAC 50  
|||||

QY 121 AACTCCCATGATCAGTGAGTGAATCTCTCCACACAATATGTCTGCTTA 169  
|||||  
Db 49 AACTCCCATGATCAGTGAGTGAATCTCTCCACACAATATGTCTGCTTA 1  
|||||

RESULT 12  
BI052559/c 373 bp mRNA linear EST 15-JUN-2001  
LOCUS

DEFINITION RCS-GN0282-220101-011-A04 GN0282 Homo sapiens cDNA, mRNA sequence.  
ACCESSION B1052559  
VERSION B1052559.1 GI:14460089  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,  
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,  
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,  
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare  
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and  
Simpson,A.J.J.  
TITLE Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
MEDLINE 20202663  
COMMENT Contact: Simpson A.J.J.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?rl=RCS5&lt2-RCS-GN0282-  
220101-011-A04&lt3-2001-01-22&lt4-1)  
Seq primer: puc 18 forward  
High quality sequence stop: 373.  
Location/Qualifiers  
1..373  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="GN0282"  
/dev\_stage="Adult"  
/note="Organ: placenta.normal; Vector: puc18; Site\_1: Sma1  
; Site\_2: Sma1; A mini-library was made by cloning  
products derived from ORESTES PCR (U.S. Letters Patent  
Application No. 196,716 - Ludwig Institute for Cancer  
Research) profiles into the puc 18 vector. Reverse  
transcription of tissue mRNA and cDNA amplification were  
performed under low stringency conditions."  
BASE COUNT 89 a 52 c 120 g 112 t  
ORIGIN  
Query Match 84.1%; Score 169; DB 13; Length 373;  
Best Local Similarity 100.0%; Pred. No. 1.5e-42;  
Matches 169; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 AACTTCAGCAGCAAGAAATAACACCACTCCGTTTAGTAGAGCACTCTGTTCCAACTCG 60  
|||||  
DB 169 AACTTCAGCAGCAAGAAATAACACCACTCCGTTTAGTAGAGCACTCTGTTCCAACTCG 110  
QY 61 GAAATTAACCCATACCTCAACCTCACTGCTGTAATAATTAGCAATACATACACACACC 120  
|||||  
DB 109 GAAATTAACCCATACCTCAACCTCACTGCTGTAATAATTAGCAATACATACACACACC 50  
QY 121 AACTCCCAATGCATCAGTGAGGTAACTCCTCCACACAATAATAGTCTGCC 169  
|||||  
DB 49 AACTCCCAATGCATCAGTGAGGTAACTCCTCCACACAATAATAGTCTGCC 1  
RESULT 13 373 bp mRNA linear EST 15-JUN-2001  
LOCUS B1052567/c  
DEFINITION RCS-GN0282-220101-011-E06 GN0282 Homo sapiens cDNA, mRNA sequence.  
ACCESSION B1052567  
VERSION B1052567.1 GI:14460097  
KEYWORDS EST.

SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,  
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,  
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,  
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare  
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and  
Simpson,A.J.J.  
TITLE Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
MEDLINE 20202663  
COMMENT Contact: Simpson A.J.J.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?rl=RCS5&lt2-RCS-GN0282-  
220101-011-E06&lt3-2001-01-22&lt4-1)  
Seq primer: puc 18 forward  
High quality sequence stop: 373.  
Location/Qualifiers  
1..373  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="GN0282"  
/dev\_stage="Adult"  
/note="Organ: placenta.normal; Vector: puc18; Site\_1: Sma1  
; Site\_2: Sma1; A mini-library was made by cloning  
products derived from ORESTES PCR (U.S. Letters Patent  
Application No. 196,716 - Ludwig Institute for Cancer  
Research) profiles into the puc 18 vector. Reverse  
transcription of tissue mRNA and cDNA amplification were  
performed under low stringency conditions."  
BASE COUNT 89 a 52 c 120 g 112 t  
ORIGIN  
Query Match 84.1%; Score 169; DB 13; Length 373;  
Best Local Similarity 100.0%; Pred. No. 1.5e-42;  
Matches 169; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 AACTTCAGCAGCAAGAAATAACACCACTCCGTTTAGTAGAGCACTCTGTTCCAACTCG 60  
|||||  
DB 169 AACTTCAGCAGCAAGAAATAACACCACTCCGTTTAGTAGAGCACTCTGTTCCAACTCG 110  
QY 61 GAAATTAACCCATACCTCAACCTCACTGCTGTAATAATTAGCAATACATACACACACC 120  
|||||  
DB 109 GAAATTAACCCATACCTCAACCTCACTGCTGTAATAATTAGCAATACATACACACACC 50  
QY 121 AACTCCCAATGCATCAGTGAGGTAACTCCTCCACACAATAATAGTCTGCC 169  
|||||  
DB 49 AACTCCCAATGCATCAGTGAGGTAACTCCTCCACACAATAATAGTCTGCC 1  
RESULT 14 379 bp mRNA linear EST 15-JUN-2001  
LOCUS B1052566/c  
DEFINITION RCS-GN0282-220101-011-E05 GN0282 Homo sapiens cDNA, mRNA sequence.  
ACCESSION B1052566  
VERSION B1052566.1 GI:14460096  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE  
AUTHORS  
1 (bases 1 to 379)  
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,  
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,  
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H.,  
Brunstein, A., deoliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,  
M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and  
Simpson, A.J.  
TITLE  
Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags  
JOURNAL  
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
COMMENT  
Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?l1-RC5&t2-RC5-GN0282-  
220101-011-E05&t3-2001-01-22&t4-1)  
Seq primer: puc 18 forward  
High quality sequence stop: 363.  
Location/Qualifiers  
1..379  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_id="GN0282"  
/dev\_stage="Adult"  
/note="Organ: placenta.normal; Vector: puc18; Site\_1: SmaI  
/ Site\_2: SmaI; A mini-library was made by cloning  
products derived from ORESTES PCR (U.S. Letters Patent  
Application No. 196,716 - Ludwig Institute for Cancer  
Research) profiles into the puc 18 vector. Reverse  
transcription of tissue mRNA and cDNA amplification were  
performed under low stringency conditions."  
BASE COUNT  
90 a 53 c 123 g 112 t 1 others  
ORIGIN  
Query Match 84.1%; Score 169; DB 13; Length 379;  
Best Local Similarity 100.0%; Pred. No. 1.5e-42;  
Matches 169; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 AACTCAGCAGCAAGAAATTAACACACCTTCCTTTAGTAGACCTCTGTTCCAAATCTG 60  
|||  
DB 169 AACTCAGCAGCAAGAAATTAACACACCTTCCTTTAGTAGACCTCTGTTCCAAATCTG 110  
QY 61 GAATAACCATACCTCAACCTCAGCTGCTGTAATAATTGACATCTACTACATACACACC 120  
|||  
DB 109 GAATAACCATACCTCAACCTCAGCTGCTGTAATAATTGACATCTACTACATACACACC 50  
QY 121 AACTCCCAATGATCAGGTGGGTAACTCTCCACACAAATAGTCTGCC 169  
|||  
DB 49 AACTCCCAATGATCAGGTGGGTAACTCTCCACACAAATAGTCTGCC 1  
RESULT 15  
BIOS2569/c 382 bp mRNA linear EST 15-JUN-2001  
LOCUS RC5-GN0282-220101-011-G04 GN0282 Homo sapiens cDNA, mRNA sequence.  
DEFINITION  
ACCESSION  
BIOS2569.1 GI:14460099  
VERSION  
KEYWORDS  
EST.  
SOURCE  
human.  
ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE  
1 (bases 1 to 382)  
Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.R.,  
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,  
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H.,

Brunstein, A., deoliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,  
M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and  
Simpson, A.J.  
TITLE  
Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags  
JOURNAL  
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
COMMENT  
Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?l1-RC5&t2-RC5-GN0282-  
220101-011-G04&t3-2001-01-22&t4-1)  
Seq primer: puc 18 forward  
High quality sequence stop: 381.  
Location/Qualifiers  
1..382  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_id="GN0282"  
/dev\_stage="Adult"  
/note="Organ: placenta.normal; Vector: puc18; Site\_1: SmaI  
/ Site\_2: SmaI; A mini-library was made by cloning  
products derived from ORESTES PCR (U.S. Letters Patent  
Application No. 196,716 - Ludwig Institute for Cancer  
Research) profiles into the puc 18 vector. Reverse  
transcription of tissue mRNA and cDNA amplification were  
performed under low stringency conditions."  
BASE COUNT  
92 a 55 c 121 g 114 t  
ORIGIN  
Query Match 84.1%; Score 169; DB 13; Length 382;  
Best Local Similarity 100.0%; Pred. No. 1.5e-42;  
Matches 169; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 AACTCAGCAGCAAGAAATTAACACACCTTCCTTTAGTAGACCTCTGTTCCAAATCTG 60  
|||  
DB 169 AACTCAGCAGCAAGAAATTAACACACCTTCCTTTAGTAGACCTCTGTTCCAAATCTG 110  
QY 61 GAATAACCATACCTCAACCTCAGCTGCTGTAATAATTGACATCTACTACATACACACC 120  
|||  
DB 109 GAATAACCATACCTCAACCTCAGCTGCTGTAATAATTGACATCTACTACATACACACC 50  
QY 121 AACTCCCAATGATCAGGTGGGTAACTCTCCACACAAATAGTCTGCC 169  
|||  
DB 49 AACTCCCAATGATCAGGTGGGTAACTCTCCACACAAATAGTCTGCC 1  
Search completed: May 2, 2003, 16:21:32  
Job time: 1115 secs

GenCore version 5.1.5  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 2, 2003, 14:52:21 ; Search time 39.333 Seconds

(without alignments)  
1567.170 Million cell updates/sec

Title: US-09-719-554-3\_COPY\_8500\_8700

Perfect score: 201

Sequence: 1 aacttcagccagcaataaa.....tattttgtctgtgtacc 201

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued\_patents\_NA:\*  
1: /cgn2\_6/p10data/2/1na/5A.COMB.seq:\*  
2: /cgn2\_6/p10data/2/1na/5B.COMB.seq:\*  
3: /cgn2\_6/p10data/2/1na/6A.COMB.seq:\*  
4: /cgn2\_6/p10data/2/1na/6B.COMB.seq:\*  
5: /cgn2\_6/p10data/2/1na/PCrUS.COMB.seq:\*  
6: /cgn2\_6/p10data/2/1na/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	201	100.0	2946	4	US-09-175-928-3
2	29	14.4	1728	1	US-08-229-781-28
3	29	14.4	1728	1	US-08-630-918-28
4	29	14.4	1728	4	US-09-004-422-28
5	28.2	14.0	465	5	PCT-US96-05320X-10
6	27.8	13.8	1010	4	US-08-453-702B-4
7	27.8	13.8	1857	4	US-09-134-001C-332
8	27.8	13.8	2248	1	US-08-639-237-1
9	27.8	13.8	2248	1	US-08-975-405-1
10	27.8	13.8	2248	4	US-09-167-109-6
11	27.8	13.8	4343	1	US-08-377-690-2
12	27.8	13.8	4360	1	US-08-486-380-1
13	27.8	13.8	4360	2	US-08-436-080-1
14	27.8	13.8	4360	2	US-08-250-848-1
15	27.8	13.8	4360	2	US-08-438-192-1
16	27.8	13.8	4360	2	US-08-475-971-1
17	27.8	13.8	4360	3	US-09-021-203-1
18	27.8	13.8	4767	4	US-09-532-584B-1
19	27.8	13.8	8697	4	US-08-961-527-123
20	27.8	13.8	65042	4	US-09-784-316-3
21	27.8	13.8	99500	4	US-09-798-096-10
22	27.6	13.7	4222	1	US-08-470-179-73
23	27.4	13.6	1135	1	US-08-229-781-49
24	27.4	13.6	1135	1	US-08-630-918-49
25	27.4	13.6	1135	1	US-09-004-422-49
26	27.4	13.6	1783	1	US-08-229-781-46
27	27.4	13.6	1783	1	US-08-630-918-46

28	27.4	13.6	1783	4	US-09-004-422-46	Sequence 46, Appl
29	27.4	13.6	1947	3	US-08-604-991-1	Sequence 1, Appl
30	27.4	13.6	1947	3	US-09-363-639-1	Sequence 1, Appl
31	27	13.4	1569	4	US-09-134-001C-918	Sequence 918, App
32	27	13.4	7244	4	US-08-378-313-26	Sequence 26, Appl
33	26.8	13.3	394	1	US-08-229-781-36	Sequence 36, Appl
34	26.8	13.3	394	1	US-08-630-918-36	Sequence 36, Appl
35	26.8	13.3	394	4	US-09-004-422-36	Sequence 36, Appl
36	26.8	13.3	730	1	US-07-826-945A-10	Sequence 10, Appl
37	26.8	13.3	1000	1	US-07-960-112B-1	Sequence 1, Appl
38	26.8	13.3	1000	1	US-07-960-112B-3	Sequence 3, Appl
39	26.8	13.3	1000	1	US-08-301-316B-1	Sequence 3, Appl
40	26.8	13.3	1000	1	US-08-301-316B-3	Sequence 3, Appl
41	26.8	13.3	1000	2	US-08-473-389B-1	Sequence 1, Appl
42	26.8	13.3	1000	2	US-08-473-389B-3	Sequence 3, Appl
43	26.8	13.3	1000	3	US-08-853-831-1	Sequence 1, Appl
44	26.8	13.3	1000	3	US-08-853-831-3	Sequence 3, Appl
45	26.8	13.3	1000	5	PCT-US93-09774-1	Sequence 1, Appl

## ALIGNMENTS

```
RESULT 1
US-09-175-928-3
: Sequence 3, Application US/09175928A
: Patent No. 6312921
: GENERAL INFORMATION:
: APPLICANT: Jacobs, Kenneth
: APPLICANT: McCoy, John M.
: APPLICANT: Lavallie, Edward R.
: APPLICANT: Collins-Racie, Lisa A.
: APPLICANT: Evans, Cheryl
: APPLICANT: Werberg, David
: APPLICANT: Treacy, Maurice
: APPLICANT: Mi, Sha
: APPLICANT: Genetics Institute, Inc.
: TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM
: FILE REFERENCE: 6006B, A1172A
: CURRENT APPLICATION NUMBER: US/09/175, 928A
: NUMBER OF FILING DATE: 1998-10-20
: NUMBER OF SEQ ID NOS: 62
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 3
: LENGTH: 2946
: TYPE: DNA
: ORGANISM: Homo sapiens
:
Query Match      100.0%: Score 201: DB 4: Length 2946:
Best Local Similarity 100.0%: Pred. No. 2.8e-52:
Matches 201: Conservative 0: Mismatches 0: Indels 0: Gaps 0:
:
QY 1 AACTTCAGCACGAAATTAACACCACTCCGTTTGTAGTAGAGCACTTGTTCATCTG 60
   |||||
DB 1549 AACTTCAGCACGAAATTAACACCACTCCGTTTGTAGTAGAGCACTTGTTCATCTG 1608
   |||||
QY 61 GAAATTAACCCATCTCAACCTCACTGCTGTAATAATTGACATACATACACAC 120
   |||||
DB 1609 GAAATTAACCCATCTCAACCTCACTGCTGTAATAATTGACATACATACACAC 1668
   |||||
QY 121 AACTCCCAATGATAGTGGGTACTCCTCCACACAAATAGTCTGCTTACCCACAGA 180
   |||||
DB 1669 AACTCCCAATGATAGTGGGTACTCCTCCACACAAATAGTCTGCTTACCCACAGA 1728
   |||||
QY 181 ATATTTTGTCTGTGTATC 201
   |||||
DB 1729 ATATTTTGTCTGTGTATC 1749
   |||||
RESULT 2
US-08-229-781-28
: Sequence 28, Application US/08229781
```

Patent No. 5589174  
GENERAL INFORMATION:  
APPLICANT: Yoshinobu OKUNO et al.  
TITLE OF INVENTION: ANTI-HUMAN INFLUENZA VIRUS ANTIBODY  
NUMBER OF SEQUENCES: 58  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Wenderoth, Lind & Ponack  
STREET: 805 Fifteenth Street, N.W., #700  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: Wordperfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/229,781  
FILING DATE: April 19, 1994  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/054,016  
FILING DATE: April 29, 1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Warren M. Cheek, Jr.  
REGISTRATION NUMBER: 33,367  
REFERENCE/DOCKET NUMBER:  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-371-8850  
TELEFAX:  
TELEX:  
INFORMATION FOR SEQ ID NO: 28:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1728 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA to genomic RNA  
HYPOTHETICAL:  
ANTI-SENSE:  
FRAGMENT TYPE:  
ORIGINAL SOURCE:  
ORGANISM: A/12um1/5/65  
STRAIN:  
INDIVIDUAL ISOLATE:  
DEVELOPMENTAL STAGE:  
HAPLOTYPE:  
TISSUE TYPE:  
CELL TYPE:  
ORGANELLE:  
IMMEDIATE SOURCE:  
LIBRARY:  
CLONE:  
POSITION IN GENOME:  
CHROMOSOME/SEGMENT:  
MAP POSITION:  
UNITS:  
FEATURE:  
NAME/KEY:  
LOCATION:  
IDENTIFICATION METHOD:  
OTHER INFORMATION:  
PUBLICATION INFORMATION:  
AUTHORS:  
TITLE:  
JOURNAL:  
VOLUME:  
ISSUE:  
PAGES:  
DATE:  
DOCUMENT NUMBER:

FILING DATE:  
PUBLICATION DATE:  
RELEVANT RESIDUES IN SEQ ID NO:  
US-08-229-781-28  
Query Match 14.4%; Score 29; DB 1; Length 1728;  
Best Local Similarity 50.4%; Pred. No. 6;  
Matches 71; Conservative 0; Mismatches 70; Indels 0; Gaps 0;  
QY 3 CTTCAGCAGGAATAACACCACTCCGTTTAGTAGACACTGTGTTCAATCGGA 62  
DB 906 CTTGGAGCAATTAATACACACTCTTTCACTGTCACACCACTGACATAGTG 965  
QY 63 AATAACCCATACCTCAACCTCAGCTGTAAATTAGCAATACATACACCA 122  
DB 966 AATGCCCAATATGTAAATCGAGAAATGTGCTTAGCAGACAGCACTAAGAAATGTT 1025  
QY 123 CTCCCAATGCATCAGTGGGT 143  
DB 1026 CCCAGATTGAATCAAGAGAT 1046  
RESULT 3  
US-08-630-918-28  
Sequence 28, Application US/08630918  
Patent No. 561350  
GENERAL INFORMATION:  
APPLICANT: Yoshinobu OKUNO et al.  
TITLE OF INVENTION: ANTI-HUMAN INFLUENZA VIRUS ANTIBODY  
NUMBER OF SEQUENCES: 58  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Wenderoth, Lind & Ponack  
STREET: 805 Fifteenth Street, N.W., #700  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20005  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: Wordperfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/630,918  
FILING DATE: April 5, 1996  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/229,781  
FILING DATE: April 19, 1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/054,016  
FILING DATE: April 29, 1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Warren M. Cheek, Jr.  
REGISTRATION NUMBER: 33,367  
REFERENCE/DOCKET NUMBER:  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-371-8850  
TELEFAX:  
TELEX:  
INFORMATION FOR SEQ ID NO: 28:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1728 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA to genomic RNA  
ORIGINAL SOURCE:  
ORGANISM: A/12um1/5/65  
US-08-630-918-28  
Query Match 14.4%; Score 29; DB 1; Length 1728;  
Best Local Similarity 50.4%; Pred. No. 6;

```

Matches 71:  Conservative  0:  Mismatches  70:  Indels  0:  Gaps  0:
OY      3  CTTGACGACAGAAATAAACACCACTTCGCGTTTGTAGAGACCTGTTGTTCCAAATCTGA  62
      |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db      906  CTTGGGGGAGCAATAATCAACACTACTCTTTTCACAAATGTCACCCACACTGACAAATAGTG  965
OY      63  AATAACCATATCCTCAACACCTGCTGTAAATTTTGCAATTACTACATACCAACCA  122
      |||  ||||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db      966  AATGCCCAAAATATGTAATAATGGAGAAATGCTGCTTAGCAACAGACTAAGAAATGTTTC  1025
OY      123  CTCCCAATGCATCAGTGGGT  143
      |  |  |  ||||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db      1026  CCCAGATTGAAATCAAGAGAT  1046

RESULT 4
US-09-004-422-28
: Sequence 28, Application US/09004422
: Patent No. 6337070
:
: GENERAL INFORMATION:
: APPLICANT: Yoshinobu OKUNO et al.
: TITLE OF INVENTION: POLYPEPTIDES FOR USE IN GENERATING
: TITLE OF INVENTION: ANTI-HUMAN INFLUENZA VIRUS ANTIBODIES (AS AMENDED)
: NUMBER OF SEQUENCES: 58
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Wenderoth, Lind & Ponack, L.L.P.
: STREET: 2033 K Street, N.W., #800
: CITY: Washington
: STATE: D.C.
: COUNTRY: U.S.A.
: ZIP: 20006
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette, 3.5 inch, 1.44 mb
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: MS-DOS
: SOFTWARE: Wordperfect 5.1
:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/004,422
: FILING DATE: January 8, 1998
: CLASSIFICATION: 530
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/443,862
: FILING DATE: May 22, 1995
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/229,781
: FILING DATE: April 19, 1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/054,016
: FILING DATE: April 29, 1993
: ATTORNEY/AGENT INFORMATION:
: NAME: Warren M. Cheek, Jr.
: REGISTRATION NUMBER: 33,367
: REFERENCE/DOCKET NUMBER:
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 202-721-8200
: TELEFAX: 202-721-8250
:
: TELEX:
: INFORMATION FOR SEQ ID NO: 28:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1728 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: linear
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA to genomic RNA
: HYPOTHEetical:
: ANTI-SENSE:
: FRAGMENT TYPE:
: ORIGINAL SOURCE:
: ORGANISM: A/12um1/5/65
: STRAIN:
: INDIVIDUAL ISOLATE:
: DEVELOPMENTAL STAGE:
: HAPLOTYPE:

```

```

TISSUE TYPE:
CELL TYPE:
CELL LINE:
ORGANELLE:
IMMEDIATE SOURCE:
LIBRARY:
CLONE:
POSITION IN GENOME:
CHROMOSOME/SEGMENT:
MAP POSITION:
UNITS:
FEATURE:
NAME/KEY:
LOCATION:
IDENTIFICATION METHOD:
OTHER INFORMATION:
PUBLICATION INFORMATION:
AUTHORS:
TITLE:
JOURNAL:
VOLUME:
ISSUE:
PAGES:
DATE:
DOCUMENT NUMBER:
FILING DATE:
PUBLICATION DATE:
RELEVANT RESIDUES IN SEQ ID NO:
US-09-004-422-28

Query Match          14.4%; Score 29; DB 4; Length 1728;
Best Local Similarity 50.4%; Fred. No. 6;
Matches 71; Conservative 0; Mismatches 70; Indels 0; Gaps 0;

OY 3 CTTGACACAGAAATAAACACCACTCCGTTTGTAGAGGACCTCTGTTCAATCTGA 62
DB 906 CTTTGAGGACATAATATACACACTACCTTTTCACATGTCACCCACTGACATAGTG 965
OY 63 AATAACCATACCTCAACCTCCTGCTGTAATAATTATACCAATACTACATACACACCA 122
DB 966 AATGCCCAATAATGTAAATGGAATGTGCTTACCAACAGSACTAAGGAATGTTG 1025
OY 123 CTCCTAATGATCAGGTGGGT 143
DB 1026 CCCAGATTGAATCAAGAGGAT 1046

RESULT 5
PCT-0596-05320A-10/c
: Sequence 10, Application PC/TUS9605320A
: GENERAL INFORMATION:
: APPLICANT: Human Genome Sciences
: APPLICANT: 9410 Key West Avenue
: APPLICANT: Rockville, MD 20850
: APPLICANT: United States of America
: APPLICANT: Johns Hopkins University
: APPLICANT: 720 Rutland Avenue
: APPLICANT: Baltimore, MD 21205
: APPLICANT: United States of America
: APPLICANT: Mark D. Adams
: APPLICANT: Owen White
: APPLICANT: Hamilton O. Smith
: APPLICANT: J. Craig Venter
: TITLE OF INVENTION: Nucleotide Sequence of the Haemophilus Influenzae Rd Genom
: NUMBER OF SEQUENCES: 48
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Sterne, Kessler, Goldstein & Fox
: STREET: 1100 New York Avenue, Suite 600
: CITY: Washington
: STATE: D.C.
: COUNTRY: USA
: ZIP: 20003-3934
: COMPUTER READABLE FORM:

```

```

MEIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/05320A
FILING DATE: April22, 1996
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/476,102
FILING DATE: June 7, 1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/487,429
FILING DATE: June 7, 1995
ATTORNEY/AGENT INFORMATION:
NAME: Eric K. Steffe
REGISTRATION NUMBER: 36,688
REFERENCE/DOCKET NUMBER: 1488.014PC01
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 371-2600
TELEFAX: (202) 371-2540
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 465 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
PCT-US96-05320A-10

```

Query Match	14.08	Score 28.2	DB 5	Length 465
Best Local Similarity	53.18	Pred. No. 7.4		
Matches 60; Conservative	0	Mismatches	53	Indels 0; Gaps 0

[illegible]

RESULT 6  
US-09-453-702B-4/c  
: Sequence 4, Application US/09453702B  
: Patent NO. 6365723  
: GENERAL INFORMATION:  
: APPLICANT: Blattner, Frederick

```

: TITLE OF INVENTION: No. 6365723e1 Sequences of E. coli 0157
:
: NUMBER OF SEQUENCES: 265
:
: CORRESPONDENCE ADDRESS:
:

```

ADDRESS: Quarles & Brady  
STREET: 1 South Pinckney Street  
CITY: Madison  
STATE: WI  
COUNTRY: US

```

; ZIP: 53701-2113
;
; COMPUTER READABLE FORM:

```

```

;
MEDIUM TYPE: Diskette, 3.50 inch. 1.44MB storage
COMPUTER: IBM PC compatible
;

```

```

;
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Word Perfect 8.0

```

```

; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/453,702B

```

; FILING DATE: 03-Dec-1999  
 ; CLASSIFICATION: <Unknown>

```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/110,955

```

FILING DATE: 04-DEC-1998

ATTORNEY/AGENT INFORMATION:  
NAME: Seay, Nicholas J.  
REGISTRATION NUMBER: 27386  
REFERENCE/DOCKET NUMBER: 960296.95017  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (608) 251-5000  
TELEFAX: (608) 251-9166  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1010  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
SEQUENCE DESCRIPTION: SEQ ID NO: 4:  
US-09-453-702B-4

SEQUENCE DESCRIPTION: SEQ ID NO: 4:  
US-09-453-702B-4

Query Match	13.8%	Score 27.8;	DB 4;	length 1010;
Best Local Similarity	54.4%;	Pred. No. 12;		
Matches	56;	Conservative	0;	Mismatches 47; Indels 0; Gaps 0;

QY    6 CAGCAGGAAATAACACACCATTCCGTTTATGATGGACCTCTGTITTCATCTGGAAAT 65  
    | | | | | | | | | | | | | | |  
DB    780 CAGGTCAGTCATACCAACGGCTATTGTTCGTTTCAGCATTCCTCAGGTCAACTCGAAGAC 721

QY	66	A	A	C	C	T	C	A	A	C	C	T	C	A	C	T	G	T	A	A	A	T	T	A	G	C	A	T	A	C	T	108								
Db	720	A	G	T	C	A	T	A	T	A	C	C	A	A	T	G	A	C	A	C	C	C	C	G	T	T	A	A	G	T	A	G	C	A	T	T	A	C	T	678

RESULT 7  
US-09-134-001C-332/c  
; sequence 332, Application US/09134001C  
Date of issue 20090520

: PATENT NO. 63803/0  
 : GENERAL INFORMATION:  
 : APPLICANT: Lynn Doucette-Stamm et al  
 : TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCCUS  
 : TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS  
 : FILE REFERENCE: GTC-007  
 : CURRENT APPLICATION NUMBER: US/09/134,001C

```

: CURRENT FILING DATE: 1998-08-13
: PRIOR APPLICATION NUMBER: US 60/064,966
: PRIOR FILING DATE: 1997-11-08
: PRIOR APPLICATION NUMBER: US 60/055,775
: PRIOR FILING DATE: 1997-08-14
: NUMBER OF SEQ. ID NOS: 5674
: SEQ. ID NO. 332

```

```
; LENGTH: 1857
; TYPE: DNA
; ORGANISM: Staphylococcus epidermidis
US-09-134-001C-332
```

Query Match	13.8%	Score 27.8;	DB 4;	Length 1857;
Best Local Similarity	53.2%	Pred. No. 14;		
Matches 59;	Conservative 0;	Mismatches 52;	Indels 0;	Gaps 0;

QY		27	TTCGGTTTACTAGSACOTCTGTTCCATCCTGGAATAACCCTACCTCAACCTCAC	86
Db		919	TTCGGTAGAAGTAAGTCGTTTAAAGCACCTAGTGATGATTCTCATCTGAAGTCCAAC	860

**QY**    87 CTGTGTAATAATTTCAGTACTACTACATACACAACCACCTCCCAATGCATCAG 137  
        |||    ||||    |||    |    ||    ||    ||    |||  
**Db**    859 CTGAGGTTAATTCCTAACTACCATCTTCAGGTAATAAACTTCATTCGATCTG 809

## RESULT 8

US-08-639-237-1/c  
; Sequence 1, Application US/08639237

; Patent No. 5710013  
; GENERAL INFORMATION:

```

; APPLICANT: Goeddel, David V.
; APPLICANT: Xiong, Jessie

```

;; TITLE OF INVENTION: No. 5710013e1 Protein - TRAF6







```
RESULT 13
US-08-436-080-1
; Sequence 1, Application US/08436080
; Patent No. 5834292
; GENERAL INFORMATION:
; APPLICANT: Rangan, Thirumale S.
; TITLE OF INVENTION: Method for Producing Somaclonal Variant
; TITLE OF INVENTION: Cotton Plants
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Christie, Parker & Hale
; STREET: P.O. Box 7068
; CITY: Pasadena
; STATE: CA
; COUNTRY: USA
; ZIP: 91109-7068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/436,080
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/122,094
; FILING DATE:
; APPLICATION NUMBER: US 07/122,200
; FILING DATE: 18-NOV-1987
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/680,048
; FILING DATE: 29-MAR-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Sharp Esq., Janice A.
; REGISTRATION NUMBER: 34,051
; REFERENCE/DOCKET NUMBER: P114:25705
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (818)795-5843
; TELEFAX: (818)577-1769
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4360 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Bacillus thuringiensis
US-08-436-080-1

Query Match          13.8%  Score 27.8;  DB 2;  Length 4360;
Best Local Similarity 49.0%;  Pred. No. 18;
Matches 74;  Conservative 0;  Mismatches 77;  Indels 0;  Gaps 0;
```

```
US-08-250-848-1
; Sequence 1, Application US/08250848
; Patent No. 5856177
; GENERAL INFORMATION:
; APPLICANT: Hudspeth, Richard L.
; APPLICANT: Grula, John W.
; TITLE OF INVENTION: PROMOTERS DERIVED FROM THE MAIZE
; TITLE OF INVENTION: PHOSPHOENOLPYRUVATE CARBOXYLASE GENE INVOLVED IN C4
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Christie, Parker & Hale
; STREET: P.O. Box 7068
; CITY: Pasadena
; STATE: CA
; COUNTRY: USA
; ZIP: 91009-7068
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentln Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/250,848
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Sharp Esq., Janice A.
; REGISTRATION NUMBER: 34,051
; REFERENCE/DOCKET NUMBER: P114:25992/JAS
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (818) 577-1769
; TELEFAX: (818) 795-5843
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4360 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Bacillus thuringiensis
US-08-250-848-1

Query Match          13.8%  Score 27.8;  DB 2;  Length 4360;
Best Local Similarity 49.0%;  Pred. No. 18;
Matches 74;  Conservative 0;  Mismatches 77;  Indels 0;  Gaps 0;
```

```
US-08-438-192-1
; Sequence 1, Application US/08438192
; Patent No. 5859321
; GENERAL INFORMATION:
; APPLICANT: Rangan, Thirumale S.
; APPLICANT: Anderson, David M.
; APPLICANT: Rajasekaran, Kanniah
; TITLE OF INVENTION: Cotton Somaclonal Variants
; NUMBER OF SEQUENCES: 1
```



GenCore version 5.1.5  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 2, 2003, 16:21:41 ; Search time 63.3333 Seconds

(Without alignments)  
3750.092 Million cell updates/sec

Title: US-09-719-554-3\_COPY\_8500\_8700

Perfect score: 201

Sequence: 1 aacttcagcagcagaataaa.....tatctttgtctgtgttacc 201

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 746064 seqs, 590810554 residues 1492128

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Maximum Match 0%

Listing first 45 summaries

Database : Published.Applications\_NA.\*  
1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*  
10: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:\*  
12: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq:\*  
13: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:\*  
14: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	201	100.0	2930	10 US-09-902-535-1	Sequence 1, Appl
2	201	100.0	2946	9 US-10-114-893-134	Sequence 134, App
3	193	96.0	1481	7 US-08-979-847-105	Sequence 105, App
4	175.4	87.3	792	10 US-09-864-761-21192	Sequence 21192, A
5	175.4	87.3	1894	10 US-09-864-761-4444	Sequence 4444, A
6	33.4	16.6	3457	10 US-09-880-107-2185	Sequence 2185, Ap
7	32.4	16.1	1084	9 US-09-938-842A-3786	Sequence 3786, Ap
8	32.4	16.1	2000	10 US-09-887-576-143	Sequence 143, App
9	30.6	15.2	3309400	9 US-09-738-626-1	Sequence 1, Appl
10	30	14.9	30	10 US-09-902-535-5	Sequence 5, Appl
11	30	14.9	580	10 US-09-822-830A-567	Sequence 567, App
12	30	14.9	1962	10 US-09-822-830A-123	Sequence 123, App
13	29.4	14.6	1024	10 US-09-925-301-15	Sequence 15, Appl
14	29.4	14.6	1999	7 US-08-781-986A-289	Sequence 289, App
15	29.4	14.6	684973	10 US-09-263-959-1	Sequence 1, Appl
16	29.2	14.5	6259	9 US-09-764-891-9689	Sequence 9689, Ap
17	29.2	14.5	6260	9 US-09-764-891-9690	Sequence 9690, Ap
18	29	14.4	1728	10 US-09-918-568-28	Sequence 28, Appl
19	29	14.4	73308	10 US-09-954-456-2276	Sequence 2276, App

C 20	29	14.4	173808	12 US-10-003-806-10	Sequence 10, Appl
C 21	28.6	14.2	659158	9 US-09-771-208-20	Sequence 20, Appl
C 22	28.4	14.1	369	10 US-09-983-965-110	Sequence 110, App
C 23	28.2	14.0	1136	10 US-09-849-967A-4	Sequence 4, Appl
C 24	28.2	14.0	1689	10 US-09-849-967A-1	Sequence 1, Appl
C 25	28	13.9	805	9 US-09-775-049-27	Sequence 27, Appl
C 26	28	13.9	958	10 US-09-864-761-9654	Sequence 9654, Ap
C 27	28	13.9	1368	9 US-09-910-186A-17	Sequence 17, Appl
C 28	27.8	13.8	1010	9 US-10-114-170-4	Sequence 4, Appl
C 29	27.8	13.8	2264	9 US-10-067-125-6	Sequence 6, Appl
C 30	27.8	13.8	4360	10 US-09-756-643-1	Sequence 1, Appl
C 31	27.8	13.8	7214	9 US-10-205-942-1	Sequence 1, Appl
C 32	27.8	13.8	34641	10 US-09-954-456-11110	Sequence 1110, Ap
C 33	27.8	13.8	34641	10 US-09-954-456-1187	Sequence 1787, Ap
C 34	27.8	13.8	65042	9 US-10-229-124-3	Sequence 3, Appl
C 35	27.6	13.7	247	10 US-09-783-590-10685	Sequence 10685, A
C 36	27.6	13.7	500	10 US-09-783-590-5616	Sequence 5616, Ap
C 37	27.4	13.6	389	9 US-09-918-568-49	Sequence 17351, A
C 38	27.4	13.6	462	10 US-09-560-863-95	Sequence 95, Appl
C 39	27.4	13.6	801	10 US-09-921-823-6	Sequence 6, Appl
C 40	27.4	13.6	1135	10 US-09-918-568-49	Sequence 49, Appl
C 41	27.4	13.6	1783	10 US-09-918-568-46	Sequence 46, Appl
C 42	27.2	13.5	342	9 US-09-933-797-156	Sequence 156, App
C 43	27.2	13.5	2867	10 US-09-954-456-1327	Sequence 1327, Ap
C 44	27.2	13.5	2867	10 US-09-880-107-2354	Sequence 2354, Ap
C 45	27.2	13.5	6005	10 US-09-070-927A-327	Sequence 327, App

## ALIGNMENTS

RESULT 1  
US-09-902-535-1  
; Sequence 1, Application US/09902535  
; Patent No. US20020102530A1  
; GENERAL INFORMATION:  
; APPLICANT: Keith, Jr., James C.  
; APPLICANT: McCoy, John M.  
; TITLE OF INVENTION: Methods and compositions for diagnosing  
; TITLE OF INVENTION: and treating preclampsia and gestational trophoblast  
; FILE REFERENCE: GIN-6006B4  
; CURRENT APPLICATION NUMBER: US/09/902,535  
; CURRENT FILING DATE: 2001-07-09  
; PRIOR APPLICATION NUMBER: 60/216,657  
; PRIOR FILING DATE: 2000-07-06  
; NUMBER OF SEQ ID NOS: 5  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1  
; LENGTH: 2930  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (930)...(2546)  
US-09-902-535-1  
Query Match 100.0%; Score 201; DB 10; Length 2930;  
Best Local Similarity 100.0%; Pred. No. 6.1e-54;  
Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 AACTTCAGCAGACGAAATTAACACGACCTGTTTACGAGACCTGTTTCAATCTG 60  
|||  
DB 1551 AACTTCAGCAGACGAAATTAACACGACCTGTTTACGAGACCTGTTTCAATCTG 1610  
OY 61 GAATTAACCATACCTCAACCTCAACCTCAACCTCAACCTCAACCTCAACCTCAAC 120  
|||  
DB 1611 GAATTAACCATACCTCAACCTCAACCTCAACCTCAACCTCAACCTCAACCTCAAC 1670  
OY 121 AACTCCCAATGATCAGTGGGTAACTCTCCACACAAATAGTCTGCTACCTCAGGA 180  
|||  
DB 1671 AACTCCCAATGATCAGTGGGTAACTCTCCACACAAATAGTCTGCTACCTCAGGA 1730

QY 181 ATATTTTGTCTGTGTACC 201  
|||||  
DB 1731 ATATTTTGTCTGTGTACC 1751

## RESULT 2

US-10-114-893-134  
Sequence 134, Application US/10114893  
Publication No. US20020193567A1  
GENERAL INFORMATION:

APPLICANT: Jacobs, Kenneth  
APPLICANT: McCoy, John M.  
APPLICANT: LaValle, Edward R.  
APPLICANT: Collins-Racie, Lisa A.  
APPLICANT: Evans, Cheryl  
APPLICANT: Merberg, David  
APPLICANT: Treacy, Maurice  
APPLICANT: Bowman, Michael R.  
APPLICANT: Spaulding, Vikki  
APPLICANT: Carlin-Duckett, McKeough  
APPLICANT: Kelleher, Kerry S.  
APPLICANT: Genetics Institute, Inc.  
TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM  
FILE REFERENCE: GI 6000-10A  
CURRENT APPLICATION NUMBER: US/10/114,893  
CURRENT FILING DATE: 2002-04-02  
EARLIER APPLICATION NUMBER: 09/413,232  
EARLIER FILING DATE: 1999-10-06  
NUMBER OF SEQ ID NOS: 321  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 134  
LENGTH: 2946  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-10-114-893-134

Query Match 100.0%; Score 201; DB 9; Length 2946;  
Best Local Similarity 100.0%; Pred. No. 6,2e-54;  
Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 AACTTCAGCAGCAATAAATACACCACTCCGTTTAGTAGAGACCTCTTCCCATCTG 60  
|||||  
DB 1549 AACTTCAGCAGCAATAAATACACCACTCCGTTTAGTAGAGACCTCTTCCCATCTG 1608  
QY 61 GAATAAACCACATACCTCAAACTCCTGTGTAAATTAGCAATACATACACAAC 120  
|||||  
DB 1609 GAATAAACCACATACCTCAAACTCCTGTGTAAATTAGCAATACATACACAAC 1668  
QY 121 AACTCCCAATGATGAGTGGGTAATCTCTCCACACAATAATAGTCTGCTTACCTCAGGA 180  
|||||  
DB 1669 AACTCCCAATGATGAGTGGGTAATCTCTCTCCACACAATAATAGTCTGCTTACCTCAGGA 1728  
QY 181 ATATTTTGTCTGTGTACC 201  
|||||  
DB 1729 ATATTTTGTCTGTGTACC 1749

## RESULT 3

US-08-979-847-105  
Sequence 105, Application US/08979847  
Publication No. US20030039664A1  
GENERAL INFORMATION:

APPLICANT: PERON, HERVE  
APPLICANT: BESEME, FREDERIC  
APPLICANT: BEDIN, FREDERIC  
APPLICANT: PARANHOS-BACCALA, GLAUCIA  
APPLICANT: KOMURIAN-PRADEL, FLORENCE  
APPLICANT: JOLIVET-REYNAUD, COLETTE  
APPLICANT: MANDRAND, BERNARD  
APPLICANT: GARSON, JEREMY  
APPLICANT: TUKE, PHILIP  
TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS

TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYL  
TITLE OF INVENTION: THERAPEUTIC PURPOSES  
NUMBER OF SEQUENCES: 210  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: OLIVE & BERRIDGE, PLC  
STREET: P.O. BOX 19928  
CITY: ALEXANDRIA  
STATE: VA  
COUNTRY: USA  
ZIP: 22320

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/979,847  
FILING DATE: 26-NOV-1997  
CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:  
NAME: BERRIDGE, WILLIAM P.  
REGISTRATION NUMBER: 30,024  
REFERENCE/DOCKET NUMBER: WPB 39046A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703-836-6400  
TELEFAX: 703-836-2787

## INFORMATION FOR SEQ ID NO: 105:

SEQUENCE CHARACTERISTICS:  
LENGTH: 1481 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
US-08-979-847-105

Query Match 96.0%; Score 193; DB 7; Length 1481;  
Best Local Similarity 97.5%; Pred. No. 1.6e-51;  
Matches 196; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 AACTTCAGCAGCAATAAATACACCACTCCGTTTAGTAGAGACCTCTTCCCATCTG 60  
|||||  
DB 622 AACTTCAGCAGCAATAAATACACCACTCCGTTTAGTAGAGACCTCTTCCCATCTG 681  
QY 61 GAATAAACCACATACCTCAAACTCCTGTGTAAATTAGCAATACATACACAAC 120  
|||||  
DB 682 GAATAAACCACATACCTCAAACTCCTGTGTAAATTAGCAATACATACACAAC 741  
QY 121 AACTCCCAATGATGAGTGGGTAATCTCTCCACACAATAATAGTCTGCTTACCTCAGGA 180  
|||||  
DB 742 AACTCCCAATGATGAGTGGGTAATCTCTCTCCACACAATAATAGTCTGCTTACCTCAGGA 801  
QY 181 ATATTTTGTCTGTGTACC 201  
|||||  
DB 802 ATATTTTGTCTGTGTACC 822

## RESULT 4

US-09-864-761-21192  
Sequence 21192, Application US/09864761  
Patent No. US20020048763A1  
GENERAL INFORMATION:

APPLICANT: Penn, Sharon G.  
APPLICANT: Rank, David R.  
APPLICANT: Hanzel, David K.  
APPLICANT: Chen, Wensheng

TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO  
TITLE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY

## FILE REFERENCE: Aecmca-x-1

CURRENT APPLICATION NUMBER: US/09/864,761  
CURRENT FILING DATE: 2001-05-23  
PRIOR APPLICATION NUMBER: US 60/180,312  
PRIOR FILING DATE: 2000-02-04  
PRIOR APPLICATION NUMBER: US 60/207,456

```

PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 21192
LENGTH: 792
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC002346.1
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 20
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 6
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 5.4
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 5.9
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 6.7
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 6.2
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 6.3
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.9
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 7.2
OTHER INFORMATION: NT HIT: AF208161.1, EVALUO 0.00e+00
OTHER INFORMATION: EST_HUMAN HIT: AU138405.1, EVALUO 0.00e+00
OTHER INFORMATION: SWISSPROT HIT: P10269, EVALUO 2.00e-16
US-09-864-761-21192

Query Match      87.3%; Score 175.4; DB 10; Length 792;
Best Local Similarity 95.5%; Pred. No. 5e-46;
Matches 192; Conservative 0; Mismatches 6; Indels 3; Gaps 1;
1
1 AACTGACAGCAAGAAATAACACCACTGCTGTTAGTAGGACCTGTTGTTCCATCAG 60
DB 248 AACTTAGACAGCAAGAAATAACACCACTGCTGTTAGTAGGCTCCTC--TTTCCAAATCTG 304
OY 61 GAAATACCACTACCTCAACCTGACCTGTTAAATTTAGCAATACATACAGCAACC 120
DB 305 GAAATACCACTACCTCAACCTGACCTGTTAAATTTAGCAATACATACAGCAACC 364
OY 121 AACTCCATGATCAGTGGGGTAACTCTCCACACAAATAGTCTGCTACCTAGGA 180
DB 365 AACTCCATGATCAGTGGGGTAACTCTCCACACAAATAGTCTGCTACCTAGGA 424
OY 181 AATATTTTGTCTGTGTAC 201
```

```

DB 425 AATATTTTGTCTGTGTAC 445
|||||
RESULT 5
US-09-864-761-4444
Sequence 4444, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO
FILE REFERENCE: Aecomica-X-1
CURRENT FILING DATE: 2001-05-23
PRIOR APPLICATION NUMBER: US 60/180,312
PRIOR FILING DATE: 2000-02-04
PRIOR APPLICATION NUMBER: US 60/207,456
PRIOR FILING DATE: 2000-05-26
PRIOR APPLICATION NUMBER: US 09/632,366
PRIOR FILING DATE: 2000-08-03
PRIOR APPLICATION NUMBER: GB 24263.6
PRIOR FILING DATE: 2000-10-04
PRIOR APPLICATION NUMBER: US 60/236,359
PRIOR FILING DATE: 2000-09-27
PRIOR APPLICATION NUMBER: PCT/US01/00666
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00667
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00664
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00669
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00665
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00668
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00663
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00662
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00661
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: PCT/US01/00670
PRIOR FILING DATE: 2001-01-30
PRIOR APPLICATION NUMBER: US 60/234,687
PRIOR FILING DATE: 2000-09-21
PRIOR APPLICATION NUMBER: US 09/608,408
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/774,203
PRIOR FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 4444
LENGTH: 1894
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC002346.1
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 20
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 6
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 5.4
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 5.9
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 4
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 6.7
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 6.2
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 6.3
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.9
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 7.2
US-09-864-761-4444
```

Query Match	87.38;	Score 175.4;	DB 10;	Length 1894;
Best Local Similarity	95.58;	Pred. No. 7.1e-46;		
Matches 192; Conservative	0;	Mismatches 6;	Indels 3;	Gaps 1.

QY	1	AACCTTCAGCAGAGAAATTAACACACAGCTCCGTTTATAGAGACCTCTGGTTCACATCTG	60
Db	446	AACCTTCAGCAGAGAAATTAACACACACAGCTCTGTTTATAGAGTCCCTC---TTTCCAAATCTG	50/20
QY	61	GAATAAACCCTATACCTCAAACTCACCCTGTGTATAAATTTAGCAATACATACACAAACC	120
Db	503	GAATTAATACCCTATACCTCAAACTCACCCTGTGTATAAATTTAGCAATACATACACAAAGCC	56/22
QY	121	AACATCCCAATGCATCAGGTGGGTAACTCTCTCCACACACAATAGTCTCTCCTACCTCAGGA	180
Db	563	AACATCCCAATGCATCAGGTGGGTAACTCTCTCCACACACAATAGTCTCTCCTACCTCAGGA	62/22
QY	181	ATATTTTGTGCTGTGGGATC	201
Db	623	ATATTTTGTGCTGTGGTACC	64/3

## RESULT 6

```

US-09-880-107-2185/c
: Sequence 2185, Application US/09880107
: Patent No. US20020142981A1
:
GENERAL INFORMATION:
:
: APPLICANT: Horne, Darci T.
: APPLICANT: Vockley, Joseph G.
: APPLICANT: Scherf, Uwe
:
: APPLICANT: Gene Logic, Inc.
:
TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer
:
FILE REFERENCE: 44921-5028-WO
:
CURRENT APPLICATION NUMBER: US/09/880.107
:
CURRENT FILING DATE: 2001-06-14
:
PRIORITY APPLICATION NUMBER: US 60/211,379
:
PRIORITY FILING DATE: 2000-06-14
:
PRIORITY APPLICATION NUMBER: US 60/237,054
:
PRIORITY FILING DATE: 2000-10-02
:
NUMBER OF SEQ ID NOS: 3950
:
SOFTWARE: PatentIn Ver. 2.1
:
SEQ ID NO 2185
:
LENGTH: 3457
:
TYPE: DNA
:
ORGANISM: Homo sapiens
:
FEATURE:
:
OTHER INFORMATION: Genbank Accession No. US20020142981A1 L26336
US-09-880-107-2185

```

Query Match	16.6%;	Score 33.4;	DB 10;	Length 3457;
Best Local Similarity	52.5%;	Pred. No. 1.4;		
Matches 73; Conservative	0;	Mismatches 66;	Indels 0;	Gaps 0;

[illegible]

## RESULT 7

US-09-938-842A-3786/c  
Sequence 3786, Application US/09938842A  
Patent No. US20020160378A1  
GENERAL INFORMATION:  
APPLICANT: Harper, Jeff  
APPLICANT: Krebs, Joel

```

: APPLICANT: Wang, Xun
: APPLICANT: Zhu, Tong
: TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
: TITLE OF INVENTION: SAME, AND METHODS OF USE
: FILE REFERENCE: SCRIPT00-3
: CURRENT APPLICATION NUMBER: US/09/938, 842A.
: CURRENT FILING DATE: 2001-08-24
: PRIOR APPLICATION NUMBER: US 60/227, 866
: PRIOR FILING DATE: 2000-08-24
: PRIOR APPLICATION NUMBER: US 60/264, 647
: PRIOR FILING DATE: 2001-01-16
: PRIOR APPLICATION NUMBER: US 60/300, 111
: PRIOR FILING DATE: 2001-06-22
: NUMBER OF SEQ ID NOS: 5379
: SEQ ID NO 3786
: LENGTH: 1084
: TYPE: DNA
: ORGANISM: Arabidopsis thaliana
: US-09-938-842A-3786

```

Query Match	16.1%;	Score 32.4;	DB 9;	Length 1084;
Best Local Similarity	56.6%;	Pred. No. 1.8;		
Matches 60;	Conservative 0;	Mismatches 46;	Indels 0;	Gaps 0;

QY 11 CAGAAATTAACACACACTCCGTTTAGTAGSAGCCTCTGTTTCCAACTGGAATTAACCC 70  
Db 576 CATAAATTAAGAAAGATGTGTACTCTTGTGAATTTTAATTCGTCTCTACTCCAAATACAA 517  
QY 71 ATACCTCAACCTCACTGTGTAAATTTAGCAATACTACTACAC 116  
Db 516 AAAACGTAAACAAACAGCTTACAAATTAATCATGAAATACATAC 471

## RESULT 8

US-09-887-576-143/c  
Sequence 143, Application us/09887576  
Patent No. US20020144047A1  
GENERAL INFORMATION:  
APPLICANT: Budworth, P.  
APPLICANT: Brown, D.  
APPLICANT: Chang, H.  
APPLICANT: Zhu, T.  
APPLICANT: Han, B.  
APPLICANT: Wang, X.  
APPLICANT: Cooper, Bret  
TITLE OF INVENTION: Promoters for regulation of plant expression  
FILE REFERENCE: 1360.001US1  
CURRENT APPLICATION NUMBER: US/09/887.576  
CURRENT FILING DATE: 2001-06-25  
PRIOR APPLICATION NUMBER: US 60/213,848  
PRIOR FILING DATE: 2000-06-23  
PRIOR APPLICATION NUMBER: US 60/214,087  
PRIOR FILING DATE: 2000-06-23  
PRIOR APPLICATION NUMBER: US 60/258,692  
PRIOR FILING DATE: 2000-12-29  
NUMBER OF SEQ ID NOS: 875  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 143  
LENGTH: 2000  
TYPE: DNA  
ORGANISM: Arabidopsis thaliana  
US-09-887-576-143

Query Match,	16.1%;	Score 32.4;	DB 10;	Length 2000;
Best Local Similarity	56.6%;	Pred. No. 2.3;		
Matches 60;	Conservative 0;	Mismatches 46;	Indels 0;	Gaps 0;

QY 11 CAGAAATTAACACCACCTCCGTTTGTGTAGAGACCTCTTGTCCAATCGGAATAACCC 70  
| | | | |  
Db 1491 CATTAATTAACACCATGTGTAGTCTTGTATTATTAATCCTATCTAATCTCCAAATACAA 1432  
| | | | |  
QY 71 ATACCTCAAACCTCACCTGTGTAATAATTAGCAATATCTATCATACAC 116  
| | | | |



Db 1431 AAAACGTAACAAACAGTTACAAATTAATCAATGAAATACAAATCAC 1386

## RESULT 9

US-09-738-626-1

; Sequence 1, Application US/09738626  
; Publication No. US20020197605A1  
; GENERAL INFORMATION:  
; APPLICANT: NAKAGAWA, SATOSHI  
; APPLICANT: MIZOGUCHI, HIROSHI  
; APPLICANT: ANDO, SEIKO  
; APPLICANT: HAYASHI, MIKIRO  
; APPLICANT: OCHIAI, KEIKO  
; APPLICANT: YOKOI, HARUHIKO  
; APPLICANT: TATEISHI, NAOKO  
; APPLICANT: SENO, AKIHIRO  
; APPLICANT: IKEDA, MASATO  
; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES  
; FILE REFERENCE: 249-125  
; CURRENT APPLICATION NUMBER: US/09/738,626  
; PRIOR FILING DATE: 2000-12-18  
; PRIOR APPLICATION NUMBER: JP 99/377484  
; PRIOR FILING DATE: 1999-12-16  
; PRIOR APPLICATION NUMBER: JP 00/159162  
; PRIOR FILING DATE: 2000-04-07  
; PRIOR APPLICATION NUMBER: JP 00/280988  
; PRIOR FILING DATE: 2000-08-03  
; NUMBER OF SEQ ID NOS: 7059  
; SOFTWARE: PatentIn ver. 3.0  
; SEQ ID NO 1  
; LENGTH: 3309400  
; TYPE: DNA  
; ORGANISM: Corynebacterium glutamicum  
US-09-738-626-1

## Query Match

Best Local Similarity 15.2%; Score 30.6; DB 9; Length 3309400;

Matches 54; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

QY 92 TAAATTTAGCATCTACATACACCAACCAATCCCAATGCAATGAGTGGTAACTCCTC 151

Db 1913847 TTTAAATTTACCAATTAATAAACAACATACACTGATTTAACTCAACTACGCC 1913906

QY 152 CCACACAAATAGTCTGCTACCCCTCAGGAATAT 184

Db 1913907 CAAACATACAAAAAGCACACCTCTCAGACATCT 1913939

## RESULT 10

US-09-902-535-5

; Sequence 5, Application US/09902535  
; Patent No. US20020102530A1  
; GENERAL INFORMATION:  
; APPLICANT: Keith, Jr., James C.  
; APPLICANT: McCoy, John M.  
; APPLICANT: M. Sha  
; TITLE OF INVENTION: Methods and compositions for diagnosing  
; TITLE OF INVENTION: and treating preclampsia and gestational trophoblast  
; FILE REFERENCE: GIN-6006B4  
; CURRENT APPLICATION NUMBER: US/09/902,535  
; CURRENT FILING DATE: 2001-07-09  
; PRIOR APPLICATION NUMBER: 60/216,657  
; PRIOR FILING DATE: 2000-07-06  
; NUMBER OF SEQ ID NOS: 5  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 5  
; LENGTH: 30  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-902-535-5

Query Match 14.9%; Score 30; DB 10; Length 30;  
Best Local Similarity 100.0%; Pred. No. 2.3;  
Matches 30; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 146 CTCCTCCACACAAATATGCTGCTACCT 175

Db 1 CTCCTCCACACAAATATGCTGCTACCT 30

## RESULT 11

US-09-822-830A-567/c

; Sequence 567, Application US/09822830A  
; Patent No. US20020142952A1  
; GENERAL INFORMATION:  
; APPLICANT: Genetics Institute, Inc.  
; APPLICANT: Wong, Gordon G.  
; APPLICANT: Clark, Hilary  
; APPLICANT: Fehchel, Kim  
; APPLICANT: Agostino, Michael J.  
; APPLICANT: Howes, Steven H.  
; APPLICANT: Resnick, Richard J.  
; APPLICANT: Gulukota, Kamalakkar  
; APPLICANT: Graham, James R.  
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS  
; FILE REFERENCE: GIN 6402  
; CURRENT APPLICATION NUMBER: US/09/822,830A  
; PRIOR FILING DATE: 2001-03-29  
; PRIOR APPLICATION NUMBER: 60/195,604  
; PRIOR FILING DATE: 2000-04-06  
; NUMBER OF SEQ ID NOS: 631  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 567  
; LENGTH: 580  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-822-830A-567

## Query Match

Best Local Similarity 14.9%; Score 30; DB 10; Length 580;

Matches 51; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

QY 33 TTTAGTAGACCTCTGTTTCCATCTGGAATACCAATCAACCTCAACCTGCTGT 92

Db 576 TTTAGACACAGACTTTGATCTATATAGATATAATCAATCAATCAATCAATCAAT 517

QY 93 AAAATTTAGCATCTACATACACAA 118

Db 516 AATATTTACAAATTAATTTATACACAA 491

## RESULT 12

US-09-822-830A-123/c

; Sequence 123, Application US/09822830A  
; Patent No. US20020142952A1  
; GENERAL INFORMATION:  
; APPLICANT: Genetics Institute, Inc.  
; APPLICANT: Wong, Gordon G.  
; APPLICANT: Clark, Hilary  
; APPLICANT: Fehchel, Kim  
; APPLICANT: Agostino, Michael J.  
; APPLICANT: Howes, Steven H.  
; APPLICANT: Resnick, Richard J.  
; APPLICANT: Gulukota, Kamalakkar  
; APPLICANT: Graham, James R.  
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS  
; FILE REFERENCE: GIN 6402  
; CURRENT APPLICATION NUMBER: US/09/822,830A  
; CURRENT FILING DATE: 2001-03-29  
; PRIOR APPLICATION NUMBER: 60/195,604  
; PRIOR FILING DATE: 2000-04-06  
; NUMBER OF SEQ ID NOS: 631  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 123

LENGTH: 1962  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-822-830A-123

Query Match 14.9%; Score 30; DB 10; Length 1962;  
Best Local Similarity 59.3%; Pred. No. 13;  
Matches 51; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

QY 33 TTATAGAGACCTCTGTTCCATGCGAATTAACCCATACCTCAACCTGCTGT 92  
DB 1202 TTATAGACAGACTTGTATCTATATTAGATATTAATCAATGCTTAATCAAAATCTAAAT 1143

QY 93 AAAATTTAGCAATACTACATACACAA 118  
DB 1142 AATATTTACAAATAATATTATACACA 1117

## RESULT 13

US-09-925-301-15/C  
Sequence 15, Application US/09925301  
Patent No. US20020052308A1  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
FILE REFERENCE: PA106  
CURRENT APPLICATION NUMBER: US/09/925,301  
CURRENT FILING DATE: 2001-08-10  
PRIOR APPLICATION NUMBER: PCT/US00/05882  
PRIOR FILING DATE: 2000-03-08  
PRIOR APPLICATION NUMBER: 60/124,270  
PRIOR FILING DATE: 1999-03-12  
NUMBER OF SEQ ID NOS: 1694  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 15  
LENGTH: 1024  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (938)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc\_feature  
LOCATION: (1005)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc\_feature  
LOCATION: (1012)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: misc\_feature  
LOCATION: (1019)  
OTHER INFORMATION: n equals a,t,g, or c  
US-09-925-301-15

Query Match 14.6%; Score 29.4; DB 10; Length 1024;  
Best Local Similarity 50.8%; Pred. No. 16;  
Matches 60; Conservative 3; Mismatches 55; Indels 0; Gaps 0;

QY 84 CACCTGTGTAATTTAGCAATACTACATACACCAACTCCCATGATAGGTGGT 143

DB 964 CACCTGTACTCGACATGTGACAGTANCAAGCAATAGCTTGACTGTGCAATGAGGT 905

QY 144 AACTCTCCACCAAAATAGTCTGCCTACCTCAGGAATATTTTGTGTGTACC 201  
DB 904 TTCTTCTCCCTTACTTAATATTCATCAATCTTCAGCAGCTTTTATCAACACTGTTCC 847

## RESULT 14

US-08-781-986A-289  
Sequence 289, Application US/08781986A  
Publication No. US20030054436A1  
GENERAL INFORMATION:  
APPLICANT: Charles Kunsch  
TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences

NUMBER OF SEQUENCES: 5255  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Human Genome Sciences, Inc.  
STREET: 9410 Key West Avenue  
CITY: Rockville  
STATE: Maryland  
COUNTRY: USA  
ZIP: 20850

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage  
COMPUTER: HP Vectra 486/33  
OPERATING SYSTEM: MSDOS version 6.2  
SOFTWARE: ASCII Text  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/781,986A  
FILING DATE:

CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Benson, Bob  
REGISTRATION NUMBER: 30,446  
REFERENCE/DOCKET NUMBER: P8248BP  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (301) 309-8504  
TELEFAX: (301) 309-8512

INFORMATION FOR SEQ ID NO: 289:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1999 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear

US-08-781-986A-289

Query Match 14.6%; Score 29.4; DB 7; Length 1999;  
Best Local Similarity 54.1%; Pred. No. 21;  
Matches 60; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 5 TCAGCAGCAATTAACACCACTCCGTTTGTAGAGACCTCTGTTCCAACTCGGAA 64  
DB 638 TTACCATGAACCTAATATTCCTATGCTCAAAATATATCTTGGCAGGAATCCAGCA 697

QY 65 TAACCATACCTCAACACCTGCTGTAATTTAGCAATACTACATACA 115  
DB 698 TATCAGTTATCTTACCTCTACATTTTAAATTAAGTAATCTCCATACA 748

RESULT 15  
US-09-263-959-1  
Sequence 1, Application US/09263959  
Patent No. US20020150891A1  
GENERAL INFORMATION:  
APPLICANT: Hood, Leroy E.  
APPLICANT: Rowen, Lee  
APPLICANT: Koop, Ben F.  
TITLE OF INVENTION: DIAGNOSTIC AND THERAPEUTIC COMPOSITIONS AND METHODS WHICH U

CORRESPONDENCE ADDRESS:  
ADDRESSEE: Seed and Berry LLP  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: US  
ZIP: 98104-7092

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/263,959  
FILING DATE: 05-MAR-1999





GenCore version 5.1.5  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 2, 2003, 12:54:45 ; Search time 689 Seconds

(Without alignments)  
8490.078 Million cell updates/sec

Title: US-09-719-554-3\_COPY\_9000\_9200

Perfect score: 201  
Sequence: 1 cctagagagagtagtgccttc.....actgagacctgggacctct 201

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

Genbmbl: \*  
1: gb.ba:\*  
2: gb.htg:\*  
3: gb.in:\*  
4: gb.om:\*  
5: gb.ov:\*  
6: gb.pat:\*  
7: gb.ph:\*  
8: gb.pl:\*  
9: gb.pr:\*  
10: gb.ro:\*  
11: gb.sts:\*  
12: gb.sy:\*  
13: gb.un:\*  
14: gb.vi:\*  
15: em.ba:\*  
16: em.fun:\*  
17: em.hum:\*  
18: em.in:\*  
19: em.mu:\*  
20: em.om:\*  
21: em.or:\*  
22: em.ov:\*  
23: em.pat:\*  
24: em.ph:\*  
25: em.pl:\*  
26: em.ro:\*  
27: em.sts:\*  
28: em.un:\*  
29: em.vi:\*  
30: em.htg.hum:\*  
31: em.htg.in:\*  
32: em.htg.other:\*  
33: em.htg.mus:\*  
34: em.htg.pln:\*  
35: em.htg.rod:\*  
36: em.htg.mam:\*  
37: em.htg.vrt:\*  
38: em.sy:\*  
39: em.htgo.hum:\*  
40: em.htgo.mus:\*  
41: em.htgo.other:\*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	201	100.0	2055	6 AX007999	AX007999 Sequence
2	201	100.0	2264	6 HSRAB7	X33499 H. sapiens m
3	201	100.0	2599	6 AX007978	AX007978 Sequence
4	201	100.0	2781	6 AX092223	AX092223 Sequence
5	201	100.0	2781	9 AF072506	AF072506 Homo sapi
6	201	100.0	2930	6 AX355872	AX355872 Sequence
7	201	100.0	2930	6 AF208161	AF208161 Homo sapi
8	201	100.0	2946	6 AK177269	AK177269 Sequence
9	201	100.0	10499	6 AX007980	AX007980 Sequence
10	201	100.0	56093	6 AX329572	AX329572 Sequence
11	201	100.0	56093	9 HSAC000064	AC000064 Human BAC
12	201	100.0	149194	9 AC007566	AC007566 Homo sapi
13	199.4	99.2	1617	9 AF156963	AF156963 Homo sapi
14	199.4	99.2	1624	9 AF506835	AF506835 Homo sapi
15	199.4	99.2	1860	9 AF513360	AF513360 Homo sapi
16	198.2	98.6	7582	6 AX000966	AX000966 Sequence
17	198.2	98.6	7582	6 AX027480	AX027480 Sequence
18	197.8	98.4	1136	6 AX000961	AX000961 Sequence
19	197.8	98.4	1136	6 AX027475	AX027475 Sequence
20	197.8	98.4	1136	6 AF072508	AF072508 Homo sapi
21	193	96.0	2782	6 AX000962	AX000962 Sequence
22	193	96.0	2782	6 AX027476	AX027476 Sequence
23	193	96.0	2782	6 AX133396	AX133396 Sequence
24	188.2	93.6	172281	6 AC068492	AC068492 Homo sapi
25	186.6	92.8	108439	2 AC019325	AC019325 Homo sapi
26	186.6	92.8	125790	9 AC106713	AC106713 Homo sapi
27	186.6	92.8	176087	2 AC024625	AC024625 Homo sapi
28	186.6	92.8	176459	2 AC108684	AC108684 Homo sapi
29	186.6	92.8	178333	9 AC108685	AC108685 Homo sapi
30	185	92.0	1684	6 AX400057	AX400057 Sequence
31	185	92.0	2074	6 AX478550	AX478550 Sequence
32	185	92.0	104853	9 AC117444	AC117444 Homo sapi
33	185	92.0	144900	9 AL391219	AL391219 Human DNA
34	185	92.0	162579	9 AL390039	AL390039 Human DNA
35	185	92.0	169462	2 AC024033	AC024033 Homo sapi
36	185	92.0	172918	2 AC023366	AC023366 Homo sapi
37	185	92.0	176305	2 AC027752	AC027752 Homo sapi
38	185	92.0	186723	2 AC092510	AC092510 Papio cyn
39	185	92.0	187321	2 AC092510	AC092510 Papio cyn
40	183.4	91.2	591	6 AX000986	AX000986 Sequence
41	183.4	91.2	591	6 AX027469	AX027469 Sequence
42	183.4	91.2	591	9 AF072498	AF072498 Homo sapi
43	183.4	91.2	1481	6 AX001027	AX001027 Sequence
44	183.4	91.2	1629	6 AX036992	AX036992 Sequence
45	183.4	91.2	1629	14 AF331500	AF331500 Multiple

#### ALIGNMENTS

RESULT 1  
LOCUS AX007999 2055 bp DNA linear PAT 06-SEP-2000  
DEFINITION Sequence 22 from Patent WO967395.  
ACCESSION AX007999  
VERSION AX007999.1 GI:9995696  
KEYWORDS  
SOURCE  
ORGANISM human.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
Perlin, J.P., Rieger, F. and Alliel, P.M.  
REFERENCE 1 (bases 1 to 2055)  
AUTHORS Nucleic sequence and deduced protein sequence family with human  
TITLE endogenous retroviral motifs, and their uses

Pred. No. is the number of results predicted by chance to have a

JOURNAL Patent: WO 9667395-A 22-29-DEC-1999;  
INST NAT SANTE RECH MED (FR); PERIN JEAN PIERRE (FR); RIEGER  
FRANCOIS (FR); ALDIEU PATRICK M (FR)  
FEATURES location/Qualifiers  
source 1. 2055  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
<1. .87  
CDS /note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAC07392.1"  
/db\_xref="GI:9995697"  
/translation="PKTANLVADITSLAKYQVLTLOGTYP"  
BASE COUNT 576 a 574 c 376 g 529 t  
ORIGIN  
Query Match 100.0%; Score 201; DB 6; Length 2055;  
Best Local Similarity 100.0%; Pred. No. 1.7e-48;  
Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 CCTAGCAGCAGTAGTCTCTCAAAATCGAAGAGCTTGAAGCTTCTAACCCTGAAGAGG 60  
DB 1512 CCTAGCAGCAGTAGTCTCTCAAAATCGAAGAGCTTGAAGCTTCTAACCCTGAAGAGG 1571  
QY 61 GGGACCTGTTATTTTATTTAGGGGAGAGATGCTGTATTTATTTATTCATCCGGAATCGT 120  
DB 1572 GGGACCTGTTATTTTATTTAGGGGAGAGATGCTGTATTTATTTATTCATCCGGAATCGT 1631  
QY 121 CACTGAGAAAGTTAAGAATTCGAGATGCAATACAGCAGAGAGAGCTTGAAA 180  
DB 1632 CACTGAGAAAGTTAAGAATTCGAGATGCAATACAGCAGAGAGAGCTTGAAA 1691  
QY 181 CACTGAGACCTGGGGCTCTCT 201  
DB 1692 CACTGAGACCTGGGGCTCTCT 1712  
RESULT 2  
HSRAB7 2264 bp mRNA linear PRI 10-FEB-1997  
LOCUS HSRAB7  
DEFINITION H. sapiens mRNA for Rab7 protein.  
ACCESSION X93499.1  
VERSION X93499.1 GI:1089892  
KEYWORDS GTP-binding protein; rab7 gene; Rab7 protein.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 2264)  
Vitel, R., Chatriello, M., Lattero, D., Brunl, C. B. and Buccil, C.  
Molecular cloning and expression analysis of the human Rab7 GTP-ase  
complementary deoxyribonucleic acid  
Biochem. Biophys. Res. Commun. 229 (3), 887-890 (1996)  
JOURNAL  
MEDLINE 97115674  
PUBMED 8954989  
REFERENCE 2 (bases 1 to 2264)  
Buccil, C.  
Direct Submission  
Submitted (27-NOV-1995) C. Buccil, University of Napoli, Dipt di  
Biologia e Patologia Cellulare, e Molecolare, II Policlinico, via  
Pansini 5, I-80131 Napoli, ITALY  
FEATURES location/Qualifiers  
source 1. 2264  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/tissue\_type="placenta"  
603. 1226  
/gene="Rab7"  
603. 1226  
/gene="Rab7"  
/function="GTP-binding protein"  
/codon\_start=1  
/product="Rab7 protein"

/protein\_id="CAA63763.1"  
/db\_xref="GI:1089893"  
/db\_xref="SWISS-PROT:P51149"  
/translation="MTSRKVLKLVILQDSGVGKTSIMNOYVKNKFSNOYKATIGAD  
FLTKREVMDDRLVTMOIDMTAGAGEFVYFGADCCVAFEPVTAAPNPKTIDSM  
RDEFLIQASPRDPENPFYVLGNKIDLEKROYATRAQAMCKSKNNIPFFETSAKEAI  
NVEQAFQITARNALKEFTEVELNEPPEPTIKDKNDRAKASAESCS"  
BASE COUNT 623 a 600 c 551 g 490 t  
ORIGIN  
Query Match 100.0%; Score 201; DB 9; Length 2264;  
Best Local Similarity 100.0%; Pred. No. 1.8e-48;  
Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 CCTAGCAGCAGTAGTCTCTCAAAATCGAAGAGCTTGAAGCTTCTAACCCTGAAGAGG 60  
DB 1848 CCTAGCAGCAGTAGTCTCTCAAAATCGAAGAGCTTGAAGCTTCTAACCCTGAAGAGG 1907  
QY 61 GGGACCTGTTATTTTATTTAGGGGAGAGATGCTGTATTTATTTATTCATCCGGAATCGT 120  
DB 1908 GGGACCTGTTATTTTATTTAGGGGAGAGATGCTGTATTTATTTATTCATCCGGAATCGT 1967  
QY 121 CACTGAGAAAGTTAAGAATTCGAGATGCAATACAGCAGAGAGAGCTTGAAA 180  
DB 1968 CACTGAGAAAGTTAAGAATTCGAGATGCAATACAGCAGAGAGAGCTTGAAA 2027  
QY 181 CACTGAGACCTGGGGCTCTCT 201  
DB 2028 CACTGAGACCTGGGGCTCTCT 2048  
RESULT 3  
AX007978 2599 bp DNA linear PAT 06-SEP-2000  
LOCUS AX007978  
DEFINITION Sequence 1 from Patent WO9967395.  
ACCESSION AX007978  
VERSION AX007978.1 GI:9995675  
KEYWORDS human.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 2599)  
Perin, J. P., Rieger, F. and Aldieu, P. M.  
Nucleic sequence and deduced protein sequence family with human  
endogenous retroviral motifs, and their uses  
Patent: WO 9667395-A 1 29-DEC-1999;  
INST NAT SANTE RECH MED (FR); PERIN JEAN PIERRE (FR); RIEGER  
FRANCOIS (FR); ALDIEU PATRICK M (FR)  
FEATURES location/Qualifiers  
source 1. 2599  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
BASE COUNT 744 a 718 c 495 g 642 t  
ORIGIN  
Query Match 100.0%; Score 201; DB 6; Length 2599;  
Best Local Similarity 100.0%; Pred. No. 1.8e-48;  
Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 CCTAGCAGCAGTAGTCTCTCAAAATCGAAGAGCTTGAAGCTTCTAACCCTGAAGAGG 60  
DB 2050 CCTAGCAGCAGTAGTCTCTCAAAATCGAAGAGCTTGAAGCTTCTAACCCTGAAGAGG 2109  
QY 61 GGGACCTGTTATTTTATTTAGGGGAGAGATGCTGTATTTATTTATTCATCCGGAATCGT 120  
DB 2110 GGGACCTGTTATTTTATTTAGGGGAGAGATGCTGTATTTATTTATTCATCCGGAATCGT 2169  
QY 121 CACTGAGAAAGTTAAGAATTCGAGATGCAATACAGCAGAGAGAGCTTGAAA 180  
DB 2170 CACTGAGAAAGTTAAGAATTCGAGATGCAATACAGCAGAGAGAGCTTGAAA 2229  
QY 181 CACTGAGACCTGGGGCTCTCT 201

|||||  
Db 2230 CACTGACCTGGGCTCT 2250

RESULT 4  
AX092223 2781 bp DNA linear PAT 21-MAR-2001  
LOCUS Sequence 2 from Patent WO0116171.  
DEFINITION AX092223  
ACCESSION AX092223  
VERSION AX092223.1 GI:134444422  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 2781)  
AUTHORS Mallet,F., Cosset,F.L., Blond,J.L., Lavillette,D., Bouton,O. and  
Ruggieri,A.  
TITLE Method for detecting the expression of an envelope protein of a  
human endogenous retrovirus and uses of a gene coding for said  
protein  
JOURNAL Patent: WO 0116171-A 2 08-MAR-2001;  
BIO MEDICAL (FR); INSTITUT NATIONAL DE LA SANTE ET DE LA RECHERCHE  
MEDICALE (INSERM) (FR)  
FEATURES  
source 1..2781  
location/Qualifiers  
BASE COUNT 747 a 770 c 556 g 708 t  
ORIGIN

Query Match 100.0%; Score 201; DB 6; Length 2781;  
Best Local Similarity 100.0%; Pred. No. 1.8e-48;  
Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCTAGCAGCAGTAGCTCTTCAAAATCGAAGAGCTTTAGACTGCTACCGCTGAAGAGG 60  
|||||  
Db 1883 CTTAGCAGCAGTAGCTCTTCAAAATCGAAGAGCTTTAGACTGCTACCGCTGAAGAGG 1942

QY 61 GGGAACTGTTATTTTATTTAGGGAAGATGCTGTTATTTATGTTATCAATCCGGATCGT 120  
|||||  
Db 1943 GGGAACTGTTATTTTATTTAGGGAAGATGCTGTTATTTATGTTATCAATCCGGATCGT 2002

QY 121 CACGAGAAAGTTAAAGAAATTCGATCGAATACACGATGAGAGGAGGCTCGAAA 180  
|||||  
Db 2003 CACTGAGAAAGTTAAAGAAATTCGATCGAATACACGATGAGAGGAGGCTCGAAA 2062

QY 181 CACTGAGACCTGGGCTCTCT 201  
|||||  
Db 2063 CACTGAGACCTGGGCTCTCT 2083

RESULT 5  
AF072506 2781 bp mRNA linear PRI 11-MAY-1999  
LOCUS Homo sapiens endogenous retrovirus W envelope protein precursor  
DEFINITION AF072506  
ACCESSION AF072506  
VERSION AF072506.2 GI:4773879  
KEYWORDS  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 2781)  
AUTHORS Blond,J.L., Beseme,F., Duret,L., Bouton,O., Bedin,F., Perron,H.,  
Mandrand,B. and Mallet,F.  
TITLE Molecular characterization and placental expression of HERV-W, a  
new human endogenous retrovirus family  
JOURNAL J. Virol. 73 (2), 1175-1185 (1999)  
MEDLINE 99099005  
PUBMED 9882319  
REFERENCE 2 (bases 1 to 2781)

AUTHORS Blond,J.-L., Beseme,F. and Mallet,F.  
TITLE Direct Submission  
JOURNAL Submitted (19-JUN-1998) UM103 CNRS-biomerieux, ENS Lyon, 46 allée  
d'Italie, Lyon, Cedex 07 69364, France  
REFERENCE 3 (bases 1 to 2781)  
AUTHORS Bouton,O., Blond,J.-L. and Mallet,F.  
TITLE Direct Submission  
JOURNAL Submitted (11-MAY-1999) UM103 CNRS-biomerieux, ENS Lyon, 46 allée  
d'Italie, Lyon, Cedex 07 69364, France  
REMARK Sequence update by submitter  
COMMENT On May 11, 1999 this sequence version replaced gi:4262291.  
FEATURES  
source 1..2781  
location/Qualifiers  
repeat\_region 1..2781  
/rpt\_family="HERV-W"  
5'UTR <1..49  
/note="R"  
5'UTR 50..503  
/note="R"  
primer\_bind 507..524  
/note="R"  
misc\_feature 534..535  
/note="putative primer binding site for tRNA-W"  
CDS 534..535  
/note="putative splice junction"  
762..2378  
/product="envelope protein precursor"  
/protein\_id="AADI4546.2"  
/db\_xref="GI:4773880"  
/translation="MALPEHLFETVLLPSTLTAPPCRCMTSSSPYQELRMQR  
GNIDAPSYRLSKGPTPTATHTMRNRCYHSATLCMHANTHTYTKMGKINPSCPGGLV  
TVCWTFFOTGSDGGGVODARREKHWVLSLTLYRGTSPPYKGLDLSKHLHRT  
HRLVSLFNTLTGLHEVSAQNPNPCICLIPNRPVYSIVPPQMNWFSREITTSV  
LVGPLVSLNLETHHSNLTVCYFSNTTYTNNQCIRWYTPPIQICLBSGIFVCGTSA  
YRCLNGSSSESCFSLFVPPMTITTEDDLSYVSKPRNKVPILIPVIGVGLGALG  
TGIGITTSYTFYKLSGLNDMEKRVADSLVTDODNSLAAYVLOMRALDLTAE  
RGTCLEFGEECCYVNOGIVPEKVEIRIDRIORAEELNTPGWLSSOMMWILP  
FLGPLAIIILLFPGCTFENLVNFSRIRAVKLQMPKQSKTKIYRPLDRASP  
RSDVNDIKGTPEELISAQPLLRNSAGSS"  
sig\_peptide 762..824  
mat\_peptide 825..2375  
misc\_feature /product="envelope protein"  
1701..1712  
/note="furin cleavage site"  
misc\_feature 1713..2375  
/note="transmembrane-region site"  
1890..1949  
/note="Region: immunosuppressive region"  
3'UTR 2347..2763  
/note="U3-R junction undetermined"  
misc\_feature 2425..2435  
/note="polypurine tract"  
polya\_site 2745..2750  
BASE COUNT 747 a 770 c 556 g 708 t  
ORIGIN

Query Match 100.0%; Score 201; DB 9; Length 2781;  
Best Local Similarity 100.0%; Pred. No. 1.8e-48;  
Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCTAGCAGCAGTAGCTCTTCAAAATCGAAGAGCTTTAGACTGCTACCGCTGAAGAGG 60  
|||||  
Db 1883 CTTAGCAGCAGTAGCTCTTCAAAATCGAAGAGCTTTAGACTGCTACCGCTGAAGAGG 1942

QY 61 GGGAACTGTTATTTTATTTAGGGAAGATGCTGTTATTTATGTTATCAATCCGGATCGT 120  
|||||  
Db 1943 GGGAACTGTTATTTTATTTAGGGAAGATGCTGTTATTTATGTTATCAATCCGGATCGT 2002

```

OY 121 CACTGAGAAAGTTAAGAAATTCGATGCAATACAGCTAGACGAGAGCTTGAAA 180
DB 2003 CACTGAGAAAGTTAAGAAATTCGATGCAATACAGCTAGACGAGAGCTTGAAA 2062
OY 181 CACTGAGACCTGGGGCTCT 201
DB 2063 CACTGAGACCTGGGGCTCT 2083

RESULT 6
AX355872 2930 bp DNA linear PAT 06-FEB-2002
LOCUS DEFINITION Sequence 1 from Patent WO0204678.
ACCESSION AX355872
VERSION AX355872.1 GI:18620523
KEYWORDS
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1
AUTHORS Keith,J.C., McCoy,J.M. and M.S.
TITLE Methods and compositions for diagnosing and treating preclampsia
and gestational trophoblast disorders
JOURNAL Patent: WO 0204678-A 1 17-JAN-2002;
GENETICS INSTITUTE, INC. (US)
FEATURES
Source location/Qualifiers
1..2930
/organism="Homo sapiens"
/db_xref="taxon:9606"
930..2546
/notes="unnamed protein product"
/codon_start=1
/protein_id="CAD23000.1"
/db_xref="GI:18620524"
/translation="MALPFIIFLFTVLLPSEFTLAPPCRCMTSSSPYOFELRMORP
GNIDAPSYRSLKGTPEFTAHMPCRYSATLCMHANHYTGKINSCGGLAY
TVCGVTFQMGSDGGGVDOAAREKHVEYSOLTRVHGTSPPKGLDLSKLTHT
HTLVLSFNTLGLHVSQNPNTNCICLPNRPVSIPIVEQNNSETEINTTSV
LVGPLVSNLEIHTHTSNLTCVKSNTTYTNSOCIRWVTPTOIVCLPSSGIFVCGTSA
YRLGNSSESMCLSFLEPMTIYTEODLYNYISKRNKRVPIIPEVIGAVYGAAG
TGIGGTTSTQFYKLSOELNGEMERADSLVLODLSLAVALVONRAIDLTAIE
RGCTCLFIEGECYVYNOGSIYEKVEIRDRIOREAEELRTNGPGLLSQMPHILP
FLGPLAIIILLFEGPCIFNLVNFVSSRIEAVKLQMEPRMOSKTKIYRPLDRPASP
RSDVNDIKGTPPEISNAQPLRPNSAGSS"
BASE COUNT 842 a 800 c 571 g 717 t
ORIGIN

Query Match 100.0%; Score 201; DB 6; Length 2930;
Best Local Similarity 100.0%; Pred. No. 1.8e-48;
Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CCTAGCAGCAGTAGCTCTTCAAAATCGAAGCTTTAGACTGCTAACCGCTGAAGAG 60
DB 2051 CCTAGCAGCAGTAGCTCTTCAAAATCGAAGCTTTAGACTGCTAACCGCTGAAGAG 2110
OY 61 GGGAACTGTTATTTTATTTAGGGGAAGATGCTGTATATGTTATATCCGAATCGT 120
DB 2111 GGGAACTGTTATTTTATTTAGGGGAAGATGCTGTATATGTTATATCCGAATCGT 2170
OY 121 CACTGAGAAAGTTAAGAAATTCGAGATCGAATACAGCTAGACGAGAGCTTGAAA 180
DB 2171 CACTGAGAAAGTTAAGAAATTCGAGATCGAATACAGCTAGACGAGAGCTTGAAA 2230
OY 181 CACTGAGACCTGGGGCTCT 201
DB 2231 CACTGAGACCTGGGGCTCT 2251

RESULT 7
AF208161 2930 bp mRNA linear PRI 22-FEB-2000
LOCUS AF208161

```

```

DEFINITION Homo sapiens syncytin precursor, mRNA, complete cds.
ACCESSION AF208161
VERSION AF208161.1 GI:6760400
KEYWORDS
SOURCE
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 2930)
AUTHORS Mi,S., Lee,X., Li,X., Veldman,G.M., Finnerty,H., Racie,L.,
Lavallee,E., Tang,X.Y., Edouard,P., Howes,S., Keith,J.C. Jr. and
McCoy,J.M.
TITLE Syncytin is a captive retroviral envelope protein involved in human
placental morphogenesis
JOURNAL Nature 403 (6771), 785-789 (2000)
MEDLINE 20155476
PUBMED 10693809
REFERENCE 2 (bases 1 to 2930)
AUTHORS Sha,M., Lee,X., Li,X., Veldman,G.M., Finnerty,H., Racie,L.,
Lavallee,E., Tang,X., Edouard,P., Howes,S., Keith,J.C. Jr. and
McCoy,J.M.
TITLE Direct Submision
JOURNAL Submitted (26-NOV-1999) Genetics Institute, 87 Cambridge Park
Drive, Cambridge, MA 02140, USA
FEATURES
Source location/Qualifiers
1..2930
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="AJ172-25"
/tissue_type="testis"
1..2930
/rpt_family="Human endogenous retrovirus HERV-W"
5'UTR 1..929
CDS 930..2546
/notes="envelope protein"
/codon_start=1
/product="syncytin precursor"
/protein_id="AAE28334.1"
/db_xref="GI:6760401"
/translation="MALPFIIFLFTVLLPSEFTLAPPCRCMTSSSPYOFELRMORP
GNIDAPSYRSLKGTPEFTAHMPCRYSATLCMHANHYTGKINSCGGLAY
TVCGVTFQMGSDGGGVDOAAREKHVEYSOLTRVHGTSPPKGLDLSKLTHT
HTLVLSFNTLGLHVSQNPNTNCICLPNRPVSIPIVEQNNSETEINTTSV
LVGPLVSNLEIHTHTSNLTCVKSNTTYTNSOCIRWVTPTOIVCLPSSGIFVCGTSA
YRLGNSSESMCLSFLEPMTIYTEODLYNYISKRNKRVPIIPEVIGAVYGAAG
TGIGGTTSTQFYKLSOELNGEMERADSLVLODLSLAVALVONRAIDLTAIE
RGCTCLFIEGECYVYNOGSIYEKVEIRDRIOREAEELRTNGPGLLSQMPHILP
FLGPLAIIILLFEGPCIFNLVNFVSSRIEAVKLQMEPRMOSKTKIYRPLDRPASP
RSDVNDIKGTPPEISNAQPLRPNSAGSS"
BASE COUNT 842 a 800 c 571 g 717 t
ORIGIN

Query Match 100.0%; Score 201; DB 9; Length 2930;
Best Local Similarity 100.0%; Pred. No. 1.8e-48;
Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CCTAGCAGCAGTAGCTCTTCAAAATCGAAGCTTTAGACTGCTAACCGCTGAAGAG 60
DB 2051 CCTAGCAGCAGTAGCTCTTCAAAATCGAAGCTTTAGACTGCTAACCGCTGAAGAG 2110
OY 61 GGGAACTGTTATTTTATTTAGGGGAAGATGCTGTATATGTTATATCCGAATCGT 120

```



Db 2111 GGGACCTGTTATTTTAAAGGAGAGATGCTGTTATTAATCAATCCGGAATCGT 2170  
QY 121 CACTGAGAAAGTTAAAGAAATTCAGATCGAATACAGAGAGAGAGCTTCGAAA 180  
Db 2171 CACTGAGAAAGTTAAAGAAATTCAGATCGAATACAGAGAGAGAGCTTCGAAA 2230  
QY 181 CACTGAGACCTGGGGCCTCCT 201  
Db 2231 CACTGAGACCTGGGGCCTCCT 2251

RESULT 8  
AR177269  
LOCUS AR177269 2946 bp DNA linear PAT 17-DEC-2001  
DEFINITION Sequence 3 from patent US 6312921.  
ACCESSION AR177269  
VERSION AR177269.1 GI:17919624  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 2946)  
AUTHORS Jacobs,K., McCoy,J.M., LaVallie,E.R., Racle,L.A., Evans,C.,  
Merberg,D., M.S. and Treacy,M.  
TITLE Secreted proteins and polynucleotides encoding them.  
JOURNAL Patent: US 6312921-A 3 06-NOV-2001;  
FEATURES  
source 1.2946  
location/Qualifiers  
BASE COUNT 858 a 801 c 570 g 717 t  
ORIGIN

Query Match 100.0%; Score 201; DB 6; Length 2946;  
Best Local Similarity 100.0%; Pred. No. 1.8e-48;  
Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCTAGACAGCAGTAGCTCCTCAAAATCGAAGAGCTTTAGACTGCTAACCGCTGAAGAGG 60  
Db 2049 CCTAGACAGCAGTAGCTCCTCAAAATCGAAGAGCTTTAGACTGCTAACCGCTGAAGAGG 2108  
QY 61 GGGAACTGTTATTTTAAAGGAGAAATGCTGTTATTAATCAATCCGGAATCGT 120  
Db 2109 GGGAACTGTTATTTTAAAGGAGAAATGCTGTTATTAATCAATCCGGAATCGT 2168  
QY 121 CACTGAGAAAGTTAAAGAAATTCAGATCGAATACAGAGAGAGAGCTTCGAAA 180  
Db 2169 CACTGAGAAAGTTAAAGAAATTCAGATCGAATACAGAGAGAGAGCTTCGAAA 2228  
QY 181 CACTGAGACCTGGGGCCTCCT 201  
Db 2229 CACTGAGACCTGGGGCCTCCT 2249

RESULT 9  
AX007980  
LOCUS AX007980 10499 bp DNA linear PAT 06-SEP-2000  
DEFINITION Sequence 3 from Patent W09967395.  
ACCESSION AX007980  
VERSION AX007980.1 GI:9995677  
KEYWORDS human.  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
REFERENCE 1 (bases 1 to 10499)  
AUTHORS Perin,J.P., Rieger,F. and Alliel,P.M.  
TITLE Nucleic sequence and deduced protein sequence family with human  
endogenous retroviral motifs, and their uses  
JOURNAL INST NAT SANTE RECH MED (FR); PERIN JEAN PIERRE (FR); RIEGER  
FRANCOIS (FR); ALLIEL PATRICK M (FR)  
FEATURES  
source 1.10499  
location/Qualifiers

BASE COUNT 3048 a 2676 c 2280 g 2495 t  
ORIGIN

Query Match 100.0%; Score 201; DB 6; Length 10499;  
Best Local Similarity 100.0%; Pred. No. 1.9e-48;  
Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCTAGACAGCAGTAGCTCCTCAAAATCGAAGAGCTTTAGACTGCTAACCGCTGAAGAGG 60  
Db 9000 CCTAGACAGCAGTAGCTCCTCAAAATCGAAGAGCTTTAGACTGCTAACCGCTGAAGAGG 9059  
QY 61 GGGAACTGTTATTTTAAAGGAGAAATGCTGTTATTAATCAATCCGGAATCGT 120  
Db 9060 GGGAACTGTTATTTTAAAGGAGAAATGCTGTTATTAATCAATCCGGAATCGT 9119  
QY 121 CACTGAGAAAGTTAAAGAAATTCAGATCGAATACAGAGAGAGAGCTTCGAAA 180  
Db 9120 CACTGAGAAAGTTAAAGAAATTCAGATCGAATACAGAGAGAGAGCTTCGAAA 9179  
QY 181 CACTGAGACCTGGGGCCTCCT 201  
Db 9180 CACTGAGACCTGGGGCCTCCT 9200

RESULT 10  
AX329572  
LOCUS AX329572 56093 bp DNA linear PAT 09-JAN-2002  
DEFINITION Sequence 81 from Patent W00194629.  
ACCESSION AX329572  
VERSION AX329572.1 GI:18102550  
KEYWORDS human.  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
REFERENCE 1  
AUTHORS Young,P.E., Augustus,M., Carter,K.C., Ebner,R., Endress,G.,  
Horligan,S., Soppet,D.R. and Weaver,Z.  
TITLE Cancer gene determination and therapeutic screening using signature  
JOURNAL Patent: WO 0194629-A 81 13-DEC-2001;  
Avalon Pharmaceuticals (US)  
FEATURES  
source 1.56093  
location/Qualifiers  
BASE COUNT 16164 a 12346 c 10702 g 16881 t  
ORIGIN

Query Match 100.0%; Score 201; DB 6; Length 56093;  
Best Local Similarity 100.0%; Pred. No. 2.1e-48;  
Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCTAGACAGCAGTAGCTCCTCAAAATCGAAGAGCTTTAGACTGCTAACCGCTGAAGAGG 60  
Db 37000 CCTAGACAGCAGTAGCTCCTCAAAATCGAAGAGCTTTAGACTGCTAACCGCTGAAGAGG 37059  
QY 61 GGGAACTGTTATTTTAAAGGAGAAATGCTGTTATTAATCAATCCGGAATCGT 120  
Db 37060 GGGAACTGTTATTTTAAAGGAGAAATGCTGTTATTAATCAATCCGGAATCGT 37119  
QY 121 CACTGAGAAAGTTAAAGAAATTCAGATCGAATACAGAGAGAGAGCTTCGAAA 180  
Db 37120 CACTGAGAAAGTTAAAGAAATTCAGATCGAATACAGAGAGAGAGCTTCGAAA 37179  
QY 181 CACTGAGACCTGGGGCCTCCT 201  
Db 37180 CACTGAGACCTGGGGCCTCCT 37200

RESULT 11

HSAC000064 56093 bp DNA linear PRI 13-NOV-1996  
 LOCUS HSAC000064  
 DEFINITION Human BAC clone RG083M05 from 7q21-7q22, complete sequence.  
 ACCESSION AC000064  
 VERSION AC000064.1 GI:1669369  
 KEYWORDS HTG.  
 SOURCE Homo sapiens.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 REFERENCE  
 1 (bases 1 to 56093)  
 2 (bases 1 to 56093)  
 3 (bases 1 to 56093)  
 4 (bases 1 to 56093)  
 5 (bases 1 to 56093)  
 6 (bases 1 to 56093)  
 7 (bases 1 to 56093)  
 8 (bases 1 to 56093)  
 9 (bases 1 to 56093)  
 10 (bases 1 to 56093)  
 11 (bases 1 to 56093)  
 12 (bases 1 to 56093)  
 13 (bases 1 to 56093)  
 14 (bases 1 to 56093)  
 15 (bases 1 to 56093)  
 16 (bases 1 to 56093)  
 17 (bases 1 to 56093)  
 18 (bases 1 to 56093)  
 19 (bases 1 to 56093)  
 20 (bases 1 to 56093)  
 21 (bases 1 to 56093)  
 22 (bases 1 to 56093)  
 23 (bases 1 to 56093)  
 24 (bases 1 to 56093)  
 25 (bases 1 to 56093)  
 26 (bases 1 to 56093)  
 27 (bases 1 to 56093)  
 28 (bases 1 to 56093)  
 29 (bases 1 to 56093)  
 30 (bases 1 to 56093)  
 31 (bases 1 to 56093)  
 32 (bases 1 to 56093)  
 33 (bases 1 to 56093)  
 34 (bases 1 to 56093)  
 35 (bases 1 to 56093)  
 36 (bases 1 to 56093)  
 37 (bases 1 to 56093)  
 38 (bases 1 to 56093)  
 39 (bases 1 to 56093)  
 40 (bases 1 to 56093)  
 41 (bases 1 to 56093)  
 42 (bases 1 to 56093)  
 43 (bases 1 to 56093)  
 44 (bases 1 to 56093)  
 45 (bases 1 to 56093)  
 46 (bases 1 to 56093)  
 47 (bases 1 to 56093)  
 48 (bases 1 to 56093)  
 49 (bases 1 to 56093)  
 50 (bases 1 to 56093)  
 51 (bases 1 to 56093)  
 52 (bases 1 to 56093)  
 53 (bases 1 to 56093)  
 54 (bases 1 to 56093)  
 55 (bases 1 to 56093)  
 56 (bases 1 to 56093)  
 57 (bases 1 to 56093)  
 58 (bases 1 to 56093)  
 59 (bases 1 to 56093)  
 60 (bases 1 to 56093)  
 61 (bases 1 to 56093)  
 62 (bases 1 to 56093)  
 63 (bases 1 to 56093)  
 64 (bases 1 to 56093)  
 65 (bases 1 to 56093)  
 66 (bases 1 to 56093)  
 67 (bases 1 to 56093)  
 68 (bases 1 to 56093)  
 69 (bases 1 to 56093)  
 70 (bases 1 to 56093)  
 71 (bases 1 to 56093)  
 72 (bases 1 to 56093)  
 73 (bases 1 to 56093)  
 74 (bases 1 to 56093)  
 75 (bases 1 to 56093)  
 76 (bases 1 to 56093)  
 77 (bases 1 to 56093)  
 78 (bases 1 to 56093)  
 79 (bases 1 to 56093)  
 80 (bases 1 to 56093)  
 81 (bases 1 to 56093)  
 82 (bases 1 to 56093)  
 83 (bases 1 to 56093)  
 84 (bases 1 to 56093)  
 85 (bases 1 to 56093)  
 86 (bases 1 to 56093)  
 87 (bases 1 to 56093)  
 88 (bases 1 to 56093)  
 89 (bases 1 to 56093)  
 90 (bases 1 to 56093)  
 91 (bases 1 to 56093)  
 92 (bases 1 to 56093)  
 93 (bases 1 to 56093)  
 94 (bases 1 to 56093)  
 95 (bases 1 to 56093)  
 96 (bases 1 to 56093)  
 97 (bases 1 to 56093)  
 98 (bases 1 to 56093)  
 99 (bases 1 to 56093)  
 100 (bases 1 to 56093)

NOTICE: This sequence may not represent the entire insert of this  
 clone. It may be shorter because we only sequence overlapping  
 sections once, or longer because we provide a small overlap between  
 neighboring submissions.

This sequence was finished as follows unless otherwise noted:  
 all regions were double stranded or sequenced with an alternate  
 chemistry; an attempt was made to resolve all sequencing problems,  
 such as compressions and repeats; all regions were covered by  
 sequence from more than one subclone; and the assembly was  
 confirmed by restriction digest.

#### SOURCE INFORMATION:

This clone is from the first release of the human BAC library. The  
 library contains cloned DNA from a human male fibroblast cell line  
 9788K. For references see: Shizuya et al., Proc. Natl. Acad. Sci.  
 89:8794-8797 (1992); Kim et al., Genomics 34:213-218 (1996).  
 VECTOR: pBELO  
 Selection: chloramphenicol

#### NEIGHBORING SEQUENCE INFORMATION:

The orientation of this clone is unknown. Actual start of this  
 clone is at base position 1 of H\_RG083M05; actual end is at 56093  
 of H\_RG083M05

This clone contains STS SW651725.

#### FEATURES

Location/Qualifiers  
 1..56093

/organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="7"  
 /map="7q21-7q22"  
 /clone="H\_RG083M05"  
 /clone\_lib="CITB-9788K-B"  
 complement(838..1131)  
 /rpt\_family="ALU"  
 repeat\_region  
 gene  
 CDS  
 /gene="WUGSC:H\_RG083M05.1"  
 join(1360..1503,4181..4370,4587..4774,6422..6556,  
 9483..9547,11631..11773,11864..12021,13131..13286,  
 14885..14988,16349..16546,16837..16971)  
 /gene="WUGSC:H\_RG083M05.1"  
 /note="ATPase; strong similarity to peroxisome  
 biosynthesis protein PAB1 (PID:91172019); coded for by  
 human cDNA C04279 (NID:91467530)"  
 /protein\_id="AAB46346.1"  
 /db\_xref="GI:1669371"  
 /translation="KRLNIQTELEVAESEAWMQPSVLLDDLLIAGLPAPPEHEH  
 SPDAEORCEILCNVKNKIDCDINKFTDLIDQHVAKETGEGFVARDFTVLVDAHSRL

SROSISTREKLVLTTLDFOKALRGFLPASLSVNLHAKPRDLGMDKIGGLHEVOILMD  
 TOLPAKVKLKKKPKYELFANLPIROFGILYGPFGKTLGLVIARESRMNFISV  
 KPELLSKYIGASEQAVDIFLRQAPACILFEPFESILAPRGHDTGYTDVRYNO  
 LITOLDGVEIGOGYVIVLARSPLDIPALIRPGLDKCYCPDPDYVITYLESKTO  
 OMLHFIYSRIEILNLVLSADPDDVLOHVASVYDSCADIKALINMOLALHG  
 MLKMSKELPDESKFNMYRLFGSSYSEIENGSTSDLSQCSAPBSMTODLPGVP  
 GKDFLFSQPVLRVTRISQGCQELTQEQDQRLADISILKGRTRSQSGHDESMOPGPI  
 KTRLAISQSHLMTALGHRPISSEDMWNAEL"  
 complement(4948..5130)  
 /rpt\_family="ALU"  
 repeat\_region  
 complement(6581..7133)  
 /rpt\_family="L1"  
 repeat\_region  
 complement(7767..8037)  
 /rpt\_family="ALU"  
 repeat\_region  
 complement(8186..8472)  
 /rpt\_family="ALU"  
 misc\_feature  
 8473..8625  
 /gene="WUGSC:H\_RG083M05.1"  
 /note="match to human 3' EST H75782 (NID:91049794), bases  
 287-444"  
 8841..9161  
 /gene="WUGSC:H\_RG083M05.1"  
 /note="match to human 5' EST H75921 (NID:91050050), bases  
 21-348"  
 9481..9547  
 /gene="WUGSC:H\_RG083M05.1"  
 /note="match to human 5' EST N22627 (NID:91130501), bases  
 276-343"  
 complement(12612..12907)  
 /rpt\_family="ALU"  
 repeat\_region  
 13670..13793  
 /gene="WUGSC:H\_RG083M05.1"  
 /note="match to human 5' EST H41382 (NID:917434), bases  
 143-266"  
 13794..13877  
 /rpt\_family="ALU"  
 repeat\_region  
 13878..13906  
 /gene="WUGSC:H\_RG083M05.1"  
 /note="match to human 5' EST H41382 (NID:917434), bases  
 30-58"  
 13907..14104  
 /rpt\_family="ALU"  
 repeat\_region  
 complement(14110..14137)  
 /rpt\_family="L1"  
 repeat\_region  
 complement(15618..15907)  
 /rpt\_family="ALU"  
 repeat\_region  
 17227..17522  
 /rpt\_family="ALU"  
 misc\_feature  
 18667..19235  
 /note="match to human fetal brain 5' EST D61494  
 (NID:970409), bases 1-255, and to human 3' EST R07476  
 (NID:9759399)"  
 19550..19670  
 /rpt\_family="ALU"  
 repeat\_region  
 21507..37303  
 /note="similarity to various SS-RNA virus polypeptides;  
 pseudogene; region of matches and close matches to  
 multiple human ESTs, see R66740 (NID:9842257)"  
 37316..37489  
 /note="Grail prediction, score = 80"  
 /evidence="not\_experimental"  
 complement(38938..39224)  
 /rpt\_family="ALU"  
 misc\_feature  
 39225..39707  
 /note="match to multiple human ESTs, see N30113  
 (NID:91148633)"  
 39800..40085  
 /rpt\_family="ALU"  
 repeat\_region  
 complement(40247..40538)  
 /rpt\_family="ALU"  
 repeat\_region  
 complement(40632..40924)  
 /rpt\_family="ALU"  
 repeat\_region  
 complement(42283..42891)

```

repeat_region /rpl_family="ALU"
complement(45474..45613)
misc_feature /rpl_family="ALU"
complement(45614..45737)
/note="match to human 3' EST H48898 (NID:g988738), bases
129-333"
misc_feature complement(46107..47026)
/note="match to multiple human ESTs, see N81064
(NID:g1243765), H48897 (NID:g988737), and M78831
(NID:g273146)"
repeat_region complement(47027..47318)
/rpl_family="ALU"
misc_feature complement(47365..47782)
/note="match to multiple human ESTs, see W37495
(NID:g1319089)"
misc_feature 47898..48115
/note="match to human 5' EST H62306 (NID:g1015138), bases
93-368"
repeat_region complement(48116..48405)
/rpl_family="ALU"
misc_feature complement(48406..48584)
/note="match to human 3' EST N29952 (NID:g1148472), bases
290-455, and 5' EST R12730 (NID:g765806)"
repeat_region complement(48787..49405)
/rpl_family="ALU"
misc_feature complement(49406..49534)
/note="match to human 3' EST R65794 (NID:g838432), bases
309-440"
repeat_region complement(49638..49672)
/rpl_family="ALU"
misc_feature complement(49674..49890)
/note="match to human 3' EST N29952 (NID:g1148472) and 5'
EST N29938 (NID:g1148458), sequences are from opposite
ends of the same clone"
gene complement(49698..51806)
/ gene="WUGSC:H.RG083M05.2"
complement(join(49698..49888,51575..51806))
/ gene="WUGSC:H.RG083M05.2"
/note="coded for by human CDNA5 W37389 (NID:g1319205),
R65891 (NID:g838529), R65794 (NID:g838432) and R65794
(NID:g838432)"
/codon_start=1
/protein_id="AAB46345.1"
/db_xref="GI:16693370"
/translation="MEFYRPGGIIICFPGYVYQGVYSDVDEQGRYYAQIRF
IDQYCEKSAALWILPTLSSPRDQEPASTIGPEEDLPRKMEYLEFVCHAPSEIFK
SRSSPPVPTREKGYIWHVGPPTPATIKESVANHL"
exon complement(51576..51758)
/ gene="WUGSC:H.RG083M05.2"
/note="Grail prediction, score = 86"
/evidence=not-experimental
repeat_region complement(52052..52329)
/rpl_family="L1"
misc_feature 55557..55843
/note="match to human EST M79192 (NID:g273505) base 2-289"
Query Match 100.0%; Score 201; DB 9; Length 56093;
Best Local Similarity 100.0%; Pred. No. 2.1e-48;
Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Db 37180 CACTGAGACCTGGGGCCTCCT 37200
|||||
RESULT 12
AC007566/c 149194 bp DNA linear PRI 01-MAR-2002
LOCUS AC007566
DEFINITION Homo sapiens BAC clone CTB-1065 from 7q21-7q22, complete sequence.
ACCESSION AC007566
VERSION AC007566.2 GI:11181861
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 149194)
Sulston, J.E. and Waterston, R.
Toward a complete human genome sequence
Genome Res. 8 (11), 1097-1108 (1998)
99063792
PUBMED 9847074
2 (bases 1 to 149194)
Du, Z.
The sequence of Homo sapiens BAC clone CTB-1065
Unpublished (2001)
3 (bases 1 to 149194)
Waterston, R.H.
Direct Submission
Submitted (15-MAY-1999) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
4 (bases 1 to 149194)
Waterston, R.
Direct Submission
Submitted (16-NOV-2000) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
5 (bases 1 to 149194)
Waterston, R.H.
Direct Submission
Submitted (02-OCT-2000) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
6 (bases 1 to 149194)
Waterston, R.H.
Direct Submission
Submitted (03-JAN-2002) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
7 (bases 1 to 149194)
Waterston, R.H.
Direct Submission
Submitted (06-FEB-2002) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
8 (bases 1 to 149194)
Waterston, R.
Direct Submission
Submitted (01-MAR-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Nov 16, 2000 this sequence version replaced g1:4835815.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: saplens@wustl.wustl.edu
----- Summary Statistics
Center project name: H_RG010605
NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

```

This sequence was finished as follows unless otherwise noted:  
all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

#### MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see  
<http://www.nhgri.nih.gov/DIR/CTB/CHR7>, send  
<mailto:egreen@nhgri.nih.gov>, or see <http://genome.wustl.edu/gsc>

#### SOURCE INFORMATION:

Clone CTB-10G5 is from the first release of the human BAC library CTB-978SK-B. The library contains cloned DNA from the male fibroblast cell line 978SK. See: Shizuya et al., Proc. Natl. Acad. Sci. USA 89:8794-7 (1992); U-U. Kim et al., Genomics 34:213-8 (1996). This clone is available from Research Genetics, Inc. (<http://www.resgen.com>).  
VECTOR: pBelOBAC11  
Selection: chloramphenicol

#### NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP5-911H5, 200 base pair overlap. Actual start of this clone is at base position 195 of CTB-10G5; actual end is at base position 150532 of CTB-10G5.

The clone CTB-10G5 contains the entire sequence of CTB-83M5.

#### FEATURES

##### source

1..149194

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/chromosome="7"

/map="7q21-7q22"

/clone="CTB-10G5"

/clone\_11b="CTB-978SK-B"

1..1634

repeat\_region

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

misc\_feature

Query Match 100.0%; Score 201; DB 9; Length 149194;

Best Local Similarity 100.0%; Pred. No. 2.2e-48;

Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 85004 CCTACGACGAGTAGCTCTTCAAAATCGAAGACCTTACGCTGACCGCTGAAGAGG 60

DB 85004 CCTACGACGAGTAGCTCTTCAAAATCGAAGACCTTACGCTGACCGCTGAAGAGG 84945

DB 85004 CCTACGACGAGTAGCTCTTCAAAATCGAAGACCTTACGCTGACCGCTGAAGAGG 84945

DB 85004 CCTACGACGAGTAGCTCTTCAAAATCGAAGACCTTACGCTGACCGCTGAAGAGG 84945

DB 85004 CCTACGACGAGTAGCTCTTCAAAATCGAAGACCTTACGCTGACCGCTGAAGAGG 84945

DB 85004 CCTACGACGAGTAGCTCTTCAAAATCGAAGACCTTACGCTGACCGCTGAAGAGG 84945

DB 85004 CCTACGACGAGTAGCTCTTCAAAATCGAAGACCTTACGCTGACCGCTGAAGAGG 84945

DB 85004 CCTACGACGAGTAGCTCTTCAAAATCGAAGACCTTACGCTGACCGCTGAAGAGG 84945

DB 85004 CCTACGACGAGTAGCTCTTCAAAATCGAAGACCTTACGCTGACCGCTGAAGAGG 84945

DB 85004 CCTACGACGAGTAGCTCTTCAAAATCGAAGACCTTACGCTGACCGCTGAAGAGG 84945

QY 121 CACTGAGAAAGTTAAAGAAATTCGATCGAATCAACGTAGACAGAGAGCTTCGAAA 180  
|||||  
Db 84884 CACTGAGAAAGTTAAAGAAATTCGATCGAATCAACGTAGACAGAGAGCTTCGAAA 84825  
QY 181 CACTGAGACCTGGGGCCTCCT 201  
|||||  
Db 84824 CACTGAGACCTGGGGCCTCCT 84804

RESULT 13  
AF156963 1617 bp DNA linear PRI 06-JUN-2000  
AF156963  
LOCUS Homo sapiens human endogenous retrovirus W envC7-1 envelope protein  
DEFINITION (env) gene, complete cds.  
ACCESSION AF156963  
VERSION AF156963.1 GI:8272467  
KEYWORDS  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 1617)  
AUTHORS Voisset,C., Boulton,O., Bedin,F., Duret,L., Mandrand,B., Mallet,F.  
TITLE Chromosomal distribution and coding capacity of the human  
JOURNAL endogenous retrovirus HERV-W family  
MEDLINE AIDS Res. Hum. Retroviruses 16 (8), 731-740 (2000)  
PUBMED 20284713  
2 (bases 1 to 1617)  
REFERENCE Voisset,C., Boulton,O., Bedin,F., Duret,L., Mandrand,B., Mallet,F.  
AUTHORS and Paranhos-Baccala,G.  
TITLE Direct Submission  
JOURNAL Submitted (09-JUN-1999) UMRI03CNRS-biomerieux, ENS Lyon, 46 allée  
d'Italie, Lyon 69364, France  
FEATURES  
source  
1. 1617  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="7"  
/clone="envC7-1"  
/note="human/rodent monochromosomal NIGMS somatic hybrid  
panel n#2 (Coriell Institute, Camden, NJ, USA); clone  
GM/NA10791 retaining human chromosome 7"  
1. 1617  
/gene="env"  
CDS  
1. 1617  
/gene="env"  
/codon\_start=1  
/product="envelope protein"  
/protein\_id="AA074215.1"  
/db\_xref="GI:8272468"  
/translation="MALPIHIFLFTVLLPSFTLTAPPSCRMSSSPYQEFLMRQRP  
GNIDAPYRSLSKGTPPTATHTPRNCYSATLCMAHNYTGKMINPSCGLGY  
TVCWTYFTQGMDSGGVODAREKHVEYISLTTRHGTSSPYKGLDSLHETLRT  
HTRLVSLPNTTLTGLHEVSAONPNCWICPLNRPVYSIPVEQNNPSEINTTSV  
LVGPLVSNLETTHTSNLCVKSNTYTTNSQCRWTPPTQVTCISGIFVCGTSA  
YRCLNGSSSEMCFLSPVPPPTITTEDDLNYSIKRPNKRVPLIPVIGAGVGAAG  
TGIGGTTSTQFYKLSQELNGDMERADSLVTLQDOLNSIAAVLONRRALDLITAE  
RGCTCLFLGEECCYVNOGIVTEKVEIRDIORRAEBELNTPGGLSCMMWMIIP  
FLGPIAAIILLFLGPCIFNLIVFVKSRIEAVKLQMEPKMQSKTKIYRRPLDRPASP  
RSDVNDIKGTPPEISAAQPLRRNSAGSS"  
repeat\_region  
1. 1617  
/rpt\_family="human endogenous retrovirus W"  
/rpt\_type="dispersed"  
1. 18  
/gene="env"  
primer\_bind  
1597..1617  
/gene="env"  
BASE COUNT 441 a 452 c 297 g 427 t  
ORIGIN  
Query Match 99.2%; Score 199.4; DB 9; Length 1617;

Best Local Similarity 99.5%; Pred. No. 5,1e-48;  
Matches 200; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 CCTAGCAGACAGTACTGCTTCAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAGAGG 60  
|||||  
Db 1122 CCTAGCAGACAGTACTGCTTCAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAGAGG 1181

QY 61 GGGAACTGTTATTTTAGGGGAAGATGCTGTTATATGTTATCAATCCGGAATCGT 120  
|||||  
Db 1182 GGGAACTGTTATTTTAGGGGAAGATGCTGTTATATGTTATCAATCCGGAATCGT 1241

QY 121 CACTGAGAAAGTTAAAGAAATTCGATCGAATCAACGTAGACAGAGAGCTTCGAAA 180  
|||||  
Db 1242 CACTGAGAAAGTTAAAGAAATTCGATCGAATCAACGTAGACAGAGAGCTTCGAAA 1301

QY 181 CACTGAGACCTGGGGCCTCCT 201  
|||||  
Db 1302 CACTGAGACCTGGGGCCTCCT 1322

RESULT 14  
AF506835 1624 bp mRNA linear PRI 20-MAY-2002  
AF506835  
LOCUS Homo sapiens envC7-1 partial cds.  
DEFINITION  
ACCESSION AF506835  
VERSION AF506835.1 GI:20978307  
KEYWORDS  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 1624)  
AUTHORS Alliel,P.M., Perin,J.P., Pierly,R., Nussbaum,J.L., Menard,A. and  
Rieger,F.  
TITLE The HERV-W/7q family in the human genome. Potential for protein  
JOURNAL expression and gene regulation  
MEDLINE Cell. Mol. Biol. 48, 213-217 (2002)  
PUBMED 99052087  
2 (bases 1 to 1624)  
REFERENCE Alliel,P.M., Perin,J.P., Goudou,D., Bitoun,M., Robert,B. and  
AUTHORS Rieger,F.  
TITLE The HERV-W/7q family in the human genome. Potential for protein  
JOURNAL expression and gene regulation  
MEDLINE Cell. Mol. Biol. 48, 213-217 (2002)  
PUBMED 99052087  
2 (bases 1 to 1624)  
REFERENCE Goudou,D., Perin,J.P., Rieger,F., Robert,B. and Alliel,P.M.  
AUTHORS Direct Submission  
JOURNAL Submitted (29-APR-2002) U-488, INSERM, 80, rue du General Leclerc,  
Le Kremlin-Bicetre 94270, France  
FEATURES  
source  
1. 1624  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/db\_xref="GI:20978308"  
/tissue\_type="placenta"  
25..>1624  
/codon\_start=1  
/product="envC7-1"  
/protein\_id="AA033413.1"  
/db\_xref="GI:20978308"  
/translation="MALPIHIFLFTVLLPSFTLTAPPSCRMSSSPYQEFLMRQRP  
GNIDAPYRSLSKGTPPTATHTPRNCYSATLCMAHNYTGKMINPSCGLGY  
TVCWTYFTQGMDSGGVODAREKHVEYISLTTRHGTSSPYKGLDSLHETLRT  
HTRLVSLPNTTLTGLHEVSAONPNCWICPLNRPVYSIPVEQNNPSEINTTSV  
LVGPLVSNLETTHTSNLCVKSNTYTTNSQCRWTPPTQVTCISGIFVCGTSA  
YRCLNGSSSEMCFLSPVPPPTITTEDDLNYSIKRPNKRVPLIPVIGAGVGAAG  
TGIGGTTSTQFYKLSQELNGDMERADSLVTLQDOLNSIAAVLONRRALDLITAE  
RGCTCLFLGEECCYVNOGIVTEKVEIRDIORRAEBELNTPGGLSCMMWMIIP  
FLGPIAAIILLFLGPCIFNLIVFVKSRIEAVKLQMEPKMQSKTKIYRRPLDRPASP  
RSDVNDIKGTPPEISAAQPLRRN"  
BASE COUNT 449 a 456 c 290 g 429 t  
ORIGIN  
Query Match 99.2%; Score 199.4; DB 9; Length 1624;

Best Local Similarity 99.5%; Pred. No. 5.1e-48;  
Matches 200; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CCTAGCAGCAGTAGTCCTTCAAAATCGAAGAGCTTTAGACTTGTACCGCTGAAGAGG 60  
|||||  
Db 1146 CTTAGCAGCAGTAGTCCTTCAAAATCGAAGAGCTTTAGACTTGTACCGCTGAAGAGG 1205

QY 61 GGGACCTGTTTATTTTATTTAGGGAAGAAATCTGTATTATTAATCAATCCGGAATCGT 120  
|||||  
Db 1206 GGGACCTGTTTATTTTATTTAGGGAAGAAATCTGTATTATTAATCAATCCGGAATCGT 1265

QY 121 CACTGAGAAAGTTAAAGAAATTCGAGATCGAATCAACAGTAGAGCAGAGAGCTTCGAAA 180  
|||||  
Db 1266 CACTGAGAAAGTTAAAGAAATTCGAGATCGAATCAACAGTAGAGCAGAGAGCTTCGAAA 1325

QY 181 CACTGACCTCGGGGCTCCT 201  
|||||  
Db 1326 CACTGACCTCGGGGCTCCT 1346

RESULT 15  
AF513360 1860 bp mRNA linear PRI 05-JUN-2002  
LOCUS Homo sapiens enverin mRNA, complete cds.  
DEFINITION AF513360  
ACCESSION AF513360.1 GI:21326140  
VERSION  
KEYWORDS  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE  
AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 1860)  
Allel, P.M., Perin, J.P., Pierig, R., Nusbaum, J.L., Menard, A. and  
Rieger, F.

TITLE Endogenous retroviruses and multiple sclerosis. Part 2: HERV-7g and  
its env transcripts  
JOURNAL C. R. Acad. Sci. III, Sci. Vie 312, 857-863 (1998)  
REFERENCE 2 (bases 1 to 1860)  
AUTHORS Allel, P.M., Perin, J.P., Goudou, D., Bitoun, M., Robert, B. and  
Rieger, F.

TITLE The HERV-W/7g family in the human genome. Potential for protein  
expression and gene regulation  
JOURNAL Cell Mol. Biol. 48 (2), 213-217 (2002)  
MEDLINE 21985840  
PUBMED 11990458

REFERENCE 3 (bases 1 to 1860)  
AUTHORS Allel, P.M., Goudou, D., Perin, J.P. and Rieger, F.  
TITLE Direct Submission  
JOURNAL Submitted (18-MAY-2002) U-488, INSERM, 80, rue du General Leclerc,  
Le Kremlin-Bicetre 94270, France

FEATURES  
source  
1. 1860  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/map="7q21"  
/issue\_type="placenta"  
59. 1675  
/codon\_start=1  
/product="enverin"  
/protein\_id="AA047599.1"  
/db\_xref="GI:21326141"  
/translation="MALPYHIFLFTVLPSPFTLAPPRCMTSSSPYOEFLMRMRP  
GNIDAPYSRLSKGPTPTFAHMPRCNCHSATLGMHANTHYWTGKMINPSCGGLGV  
TYCMTYFTQTSMDGGVODDA REKHUKVETISQLTRVHTSSPYGLDLSKHELT  
HTRVLSLENTITLTGLHEVSAQNPNCWICLPNFRVYSIPVEQMNNSTEINTSV  
LVGPLVSNLEITHSTNLVCVKSNTTYTNSOCIRWPTPTQICLPSCIFVCGTSA  
YKLGSSSSKMCFLSFLVPMPTIYTEODLYNVISKPRNKRVPILPFVIGAGVGLTG  
TGIGITTSSTOFYKLSQELNDMERVADSLVTLQDNLAAVVLONRRALDLTAE  
RGCTCLFGECCYVYVNSGIVTEKVRKEIRDRIOREBELRNTGPMGLISOMMPWILP  
FLGPLAAILILLIRGCTIFNLVNVSSSRIEAVKIQMEPKMQSKTKIYRPLDRASP  
RSDVDIKGTPEEISAOPFLRPSAGSS"

CDS

BASE COUNT 513 a 520 c 350 g 477 t

ORIGIN  
Query Match 99.2%; Score 199.4; DB 9; Length 1860;  
Best Local Similarity 99.5%; Pred. No. 5.1e-48;  
Matches 200; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CCTAGCAGCAGTAGTCCTTCAAAATCGAAGAGCTTTAGACTTGTACCGCTGAAGAGG 60  
|||||  
Db 1180 CTTAGCAGCAGTAGTCCTTCAAAATCGAAGAGCTTTAGACTTGTACCGCTGAAGAGG 1239

QY 61 GGGACCTGTTTATTTTATTTAGGGAAGAAATCTGTATTATTAATCAATCCGGAATCGT 120  
|||||  
Db 1240 GGGACCTGTTTATTTTATTTAGGGAAGAAATCTGTATTATTAATCAATCCGGAATCGT 1299

QY 121 CACTGAGAAAGTTAAAGAAATTCGAGATCGAATCAACAGTAGAGCAGAGAGCTTCGAAA 180  
|||||  
Db 1300 CACTGAGAAAGTTAAAGAAATTCGAGATCGAATCAACAGTAGAGCAGAGAGCTTCGAAA 1359

QY 181 CACTGACCTCGGGGCTCCT 201  
|||||  
Db 1360 CACTGACCTCGGGGCTCCT 1380

Search completed: May 2, 2003, 15:27:56  
Job time : 756 secs

GenCore version 5.1.5  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 2, 2003, 12:49:35 ; Search time 159.333 Seconds

(without alignments)  
2840.910 Million cell updates/sec

Title: US-09-719-554-3\_COPY\_9000\_9200

Perfect score: 201  
Sequence: 1 cctagcagcagcagcagcctc.....actgacacctggcgccctct 201

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :  
1: N.Geneseq\_101002:\*  
2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:\*  
3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:\*  
4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:\*  
5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:\*  
6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:\*  
7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:\*  
8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:\*  
9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:\*  
10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:\*  
11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:\*  
12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:\*  
13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:\*  
14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:\*  
15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:\*  
16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:\*  
17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:\*  
18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:\*  
19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:\*  
20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:\*  
21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:\*  
22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:\*  
23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:\*  
24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	201	100.0	1617	22	AAH20070
2	201	100.0	2055	21	ABN97948
3	201	100.0	2599	21	ABN97927
4	201	100.0	2781	22	AA555530
5	201	100.0	2930	24	AA24195
6	201	100.0	2946	20	AA77526
7	201	100.0	2946	21	AA59468
8	201	100.0	10499	21	ABN97929
9	201	100.0	56093	24	AB161744

10	198.2	98.6	7582	20	AA525665	Complete human end
11	198.2	98.6	7582	21	AA559215	Human endogenous r
12	197.8	98.4	1136	20	AA525660	Human endogenous r
13	197.8	98.4	1136	21	AA559210	3' pol gene and 3'
14	197.8	98.4	6394	23	AA584210	DNA encoding novel
15	193	96.0	2782	20	AA525661	Human endogenous r
16	193	96.0	2782	21	AA559211	5' non coding, 3'
17	193	96.0	2782	22	AAH20069	HERV-W envelope pr
18	183.4	91.2	591	20	AA525685	Human endogenous r
19	183.4	91.2	591	21	AA559204	Probe PenV-C15 use
20	183.4	91.2	1481	19	AA543217	Multiple sclerosis
21	183.4	91.2	1481	20	AA529703	Clone C15 from MSR
22	183.4	91.2	1629	21	AA596625	DNA encoding an en
23	183.4	91.2	2030	21	AA563826	Nucleotide sequenc
24	180.4	89.8	1088	23	AA592669	DNA encoding novel
25	180.4	89.8	1174	23	AA592498	DNA encoding novel
26	180.4	89.8	1929	23	AA571718	DNA encoding novel
27	180.4	89.8	1929	23	AA576466	DNA encoding novel
28	180.4	89.8	1929	23	AA592493	DNA encoding novel
29	180.4	89.8	1929	23	AA592666	DNA encoding novel
30	178.6	88.9	1894	22	ABA43822	Human breast cell
31	178.6	88.9	1894	22	ABA56337	Human foetal liver
32	178.6	88.9	1894	22	ABA25978	Probe #4444 for ge
33	178.6	88.9	1894	22	AAK04516	Human brain expres
34	178.6	88.9	1894	22	AAK30018	Human bone marrow
35	178.6	88.9	1894	22	AA114608	Probe #4541 for ge
36	178.6	88.9	1894	22	AA135960	Probe #4666 used t
37	178.6	88.9	1894	22	AA104422	Probe #4413 used t
38	178.6	88.9	1894	24	AB504589	Human genome-deriv
39	178.6	88.9	2784	21	ABN97930	Human retroviral s
40	164.2	81.7	1446	23	AA567605	DNA encoding novel
41	164.2	81.7	1446	23	AA573545	DNA encoding novel
42	164.2	81.7	1721	23	AA572233	DNA encoding novel
43	163.8	81.5	1329	19	AA543219	Multiple sclerosis
44	163.8	81.5	1329	20	AA529704	Clone 5M6 from MSR
45	156.2	77.7	902	23	AA571722	DNA encoding novel

## ALIGNMENTS

RESULT 1	AAH20070	standard; DNA: 1617 BP.
ID	AAH20070	
AC	AAH20070:	
XX	XX	
DE	08-AUG-2001	(first entry)
XX	XX	
DE	HERV-W envelope protein C coding sequence.	
XX	XX	
KW	Human endogenous retrovirus; HERV-W; HERV; chromosome 7; env protein;	
KW	envelope protein; multiple sclerosis-related superantigen; vaccine;	
KW	surface antigen; transmembrane; multiple sclerosis; neutrophilic;	
KW	antigen-therapy; autoimmune disorder; ds.	
XX	XX	
OS	Human endogenous retrovirus.	
XX	XX	
FT	Key	Location/Qualifiers
FT	CDS	1..1617
FT	FT	/tag- a
FT	FT	/product- "HERV-W envelope protein G"
PN	WO200131021-A1.	
PD	03-MAY-2001.	
XX	XX	
PF	30-OCT-2000; 2000WO-EP10659.	
XX	XX	
PR	28-OCT-1999; 99EP-0402690.	
XX	XX	
PA	(UYGE-) UNIV GENEVE.	
XX	XX	

PI Conrad B, Mach B;  
XX WPI: 2001-316336/33.  
DR P-PSDB; AAB75138.  
XX  
PT New human retrovirus HERV-W ENV proteins/peptides having superantigen activity useful for diagnosing and treating multiple sclerosis -  
XX  
PS Claim 13: Fig 10; 94pp; English.  
XX  
CC On the basis of the PBS t-RNA motif used for the classification of human  
CC endogenous retrovirus (HERVs) the full length endogenous provirus which  
CC was located on the long arm of human chromosome 7 (7q21-22) has been  
CC designated HERV-W. The present invention describes proteins or peptides  
CC (I) having superantigen (SAG) activity comprising the ENV protein (ENV)  
CC of HERV-W, the surface protein (SU) and transmembrane (TM) sub-units. (I)  
CC have neuroprotective activity, and can be used in: vaccines; antisense  
CC therapy; and HERV-W SAG activity-inhibitors. (I) and encoding DNA/RNA are  
CC useful for diagnosing multiple sclerosis (MS) or HERV-W-associated  
CC disorders. (I) are also useful for identifying substances (and optionally  
CC recovering) capable of binding to a retroviral superantigen associated  
CC with MS, substances capable of blocking SAG activity and substances  
CC capable of blocking transcription or translation of HERV-W retroviral  
CC superantigen. A protein or peptide derived from (I), modified to be  
CC devoid of SAG activity and being capable of generating an immune response  
CC against HERV-W retroviral SAG is useful in therapy. Nucleic acid  
CC molecules encoding (I) are useful as vaccines against MS. Substances  
CC capable of blocking SAG activity, capable of binding to a retroviral  
CC superantigen associated with MS, or capable of blocking transcription or  
CC translation of HERV-W retroviral superantigen for use in treating or  
CC preventing MS, obtained using (I) are useful for the treatment and  
CC prevention of MS. (I) and nucleic acids encoding them are useful for  
CC diagnosing autoimmune disease. The present sequence encodes the  
CC specifically claimed envelope protein of HERV-W designated G.  
XX  
SQ Sequence 1617 BP; 442 A; 452 C; 296 G; 427 T; 0 other;  
Query Match 100.0%; Score 201; DB 22; Length 1617;  
Best Local Similarity 100.0%; Pred. No. 4.9e-55;  
Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 CCTAGAGCAGTAGTCTCTCAAAATGCAAGAGCTTAGACTCTCAACCGCTGAAGAGG 60  
DB 1122 CCTAGCAGCAGTAGTCTCTCAAAATGCAAGAGCTTAGACTCTCAACCGCTGAAGAGG 1181  
QY 61 GGGACCTGTTATTTTATAGGGAGAGATGCTGTATTATGTTAATCAATCCGGAATCGT 120  
DB 1182 GGGACCTGTTATTTTATAGGGAGAGATGCTGTATTATGTTAATCAATCCGGAATCGT 1241  
QY 121 CACTGAGAAAGTTAAAGAAATTCGAGATCGAATACAGTAGAGCAGAGAGCTTGAAA 180  
DB 1242 CACTGAGAAAGTTAAAGAAATTCGAGATCGAATACAGTAGAGCAGAGAGCTTGAAA 1301  
QY 181 CACTGAGACCTGGGGCTCTCT 201  
DB 1302 CACTGAGACCTGGGGCTCTCT 1322

RESULT 2  
ABN97948  
ID ABN97948 standard; DNA; 2055 BP.  
XX  
AC ABN97948;  
XX  
DT 01-AUG-2002 (first entry)  
XX  
DE Human retroviral HERV-7q env coding sequence.  
XX  
XX Autoimmune disease; HERV-7q; chromosome 7q; immunotherapy;  
KW Multiple sclerosis; ds.  
XX  
OS Human retrovirus.  
XX

PN WO9967395-A1.  
XX  
XX 29-DEC-1999.  
PD  
XX 23-JUN-1999; 99WO-FR01513.  
PF  
XX 23-JUN-1998; 98FR-0007920.  
PR  
XX (INRM) INSERM INST NAT SANTE & RECH MEDICALE.  
XX  
XX Allel PM, Perin J, Rieger F;  
PI  
DR WPI: 2000-160587/14.  
XX  
PT New nucleic acid sequences of human endogenous retrovirus, HERV-7q,  
PT used for diagnosis, treatment and prevention of autoimmune and  
PT neurological diseases -  
PS Claim 3: Page 142-145; 225pp; French.  
XX  
CC The present invention relates to new nucleic acid sequences of human  
CC endogenous retrovirus, HERV-7q, which is located on chromosome 7q.  
CC Regulatory elements associated with HERV-7q may alter expression of other  
CC genes (even remote genes) on the same chromosome, inducing immunological  
CC and/or neurological changes (which may be pathological or protective/  
CC curative). HERV-7q peptides can be used to improve efficiency of the  
CC immune response, e.g. in immunotherapy. HERV-7q peptides and their coding  
CC sequences can be used in immunogenic or vaccinating compositions, for  
CC protection against autoimmune diseases, particularly multiple sclerosis.  
CC The peptides may also be used (by sequence comparison) to detect/identify  
CC endogenous retroviruses that are abnormally expressed in cancer,  
CC neuropathologies or other autoimmune diseases. The present sequence was  
CC used to illustrate the invention.  
XX  
SQ Sequence 2055 BP; 576 A; 574 C; 376 G; 529 T; 0 other;  
Query Match 100.0%; Score 201; DB 21; Length 2055;  
Best Local Similarity 100.0%; Pred. No. 5.4e-55;  
Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 CCTAGAGCAGTAGTCTCTCAAAATGCAAGAGCTTAGACTCTCAACCGCTGAAGAGG 60  
DB 1512 CCTAGCAGCAGTAGTCTCTCAAAATGCAAGAGCTTAGACTCTCAACCGCTGAAGAGG 1571  
QY 61 GGGACCTGTTATTTTATAGGGAGAGATGCTGTATTATGTTAATCAATCCGGAATCGT 120  
DB 1572 GGGACCTGTTATTTTATAGGGAGAGATGCTGTATTATGTTAATCAATCCGGAATCGT 1631  
QY 121 CACTGAGAAAGTTAAAGAAATTCGAGATCGAATACAGTAGAGCAGAGAGCTTGAAA 180  
DB 1632 CACTGAGAAAGTTAAAGAAATTCGAGATCGAATACAGTAGAGCAGAGAGCTTGAAA 1691  
QY 181 CACTGAGACCTGGGGCTCTCT 201  
DB 1692 CACTGAGACCTGGGGCTCTCT 1712

RESULT 3  
ABN97927  
ID ABN97927 standard; DNA; 2599 BP.  
XX  
AC ABN97927;  
XX  
DT 01-AUG-2002 (first entry)  
XX  
DE Human retroviral sequence HERV 7 env.  
XX  
XX Autoimmune disease; HERV-7q; chromosome 7q; immunotherapy;  
KW Multiple sclerosis; ds.  
XX  
OS Human retrovirus.  
XX  
PN WO9967395-A1.



```
XX 29-DEC-1999.
PD
XX 23-JUN-1999; 99WO-FR01513.
XX
PR 23-JUN-1998; 98FR-0007920.
XX
XX (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
XX
PI Alliel PM, Perin J, Rieger F;
XX
DR WPI; 2000-160587/14.
XX
PT New nucleic acid sequences of human endogenous retrovirus, HERV-7q,
PT used for diagnosis, treatment and prevention of autoimmune and
PT neurological diseases .
XX
PS Claim 1; Page 128-129; 225pp; French.
XX
CC The present invention relates to new nucleic acid sequences of human
CC endogenous retrovirus, HERV-7q, which is located on chromosome 7q.
CC Regulatory elements associated with HERV-7q may alter expression of other
CC genes (even remote genes) on the same chromosome, inducing immunological
CC and/or neurological changes (which may be pathological or protective/
CC curative). HERV-7q peptides can be used to improve efficiency of the
CC immune response, e.g. in immunotherapy. HERV-7q peptides and their coding
CC sequences can be used in immunogenic or vaccinating compositions, for
CC protection against autoimmune diseases, particularly multiple sclerosis.
CC The peptides may also be used (by sequence comparison) to detect/identify
CC endogenous retroviruses that are abnormally expressed in cancer.
CC neuropathologies or other autoimmune diseases. The present sequence was
CC used to illustrate the invention.
XX
SQ Sequence 2599 BP; 744 A; 718 C; 495 G; 642 T; 0 other;
Query Match 100.0%; Score 201; DB 21; Length 2599;
Best Local Similarity 100.0%; Pred. No. 5.8e-55;
Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CCTAGCAGCAGTAGTCTCTCAAAATCGAAGAGCTTTAGACTTGTCAACCGCTGAAGAGG 60
DB 2050 CCTAGCAGCAGTAGTCTCTCAAAATCGAAGAGCTTTAGACTTGTCAACCGCTGAAGAGG 2109
QY 61 GGGAACTGTTATTTTGGGGAAGAATGCTGTATTATTTAATCAATCCGGAATCGT 120
DB 2110 GGGAACTGTTATTTTGGGGAAGAATGCTGTATTATTTAATCAATCCGGAATCGT 2169
QY 121 CACTGAGAAAGTTAAAGAATTCGAGATCGAATACAACTAGACAGAGAGCTTCGAAA 180
DB 2170 CACTGAGAAAGTTAAAGAATTCGAGATCGAATACAACTAGACAGAGAGCTTCGAAA 2229
QY 181 CACTGAGACCTGGGGGCTCTCT 201
DB 2230 CACTGAGACCTGGGGGCTCTCT 2250
RESULT 4
AAF55630
ID AAF55630 standard; DNA; 2781 BP.
XX
AC AAF55630;
XX
DT 29-MAY-2001 (first entry)
XX
DE Nucleotide sequence of a human endogenous retrovirus envelope protein.
XX
KW Envelope protein; HERV; syncytia formation; placental development;
XX syncytia; cancer; cell adhesion; ss.
XX
OS Human endogenous retrovirus.
XX
FH Key Location/Qualifiers
FT CDS 762..2378
```

```
FT /tag= a
FT /product= "envelope protein"
FT
PN WO200116171-A1.
XX
XX 08-MAR-2001.
XX
XX 01-SEP-2000; 2000WO-FR02429.
XX
XX 01-SEP-1999; 99FR-0011141.
XX
XX 15-SEP-1999; 99FR-0011793.
XX
XX (INMR ) BIO MERIEUX.
XX
XX (INRM ) INST NAT SANTE & RECH MEDICALE.
XX
PI Mallet F, Cosset F, Blond J, Lavillette D, Boulton O, Ruggieri A;
XX
DR WPI; 2001-226676/23.
XX
DR P-PSDB; AAB67652.
XX
PT Detecting expression of human endogenous retrovirus envelope protein in
PT cells of a tissue or culture, from its ability to induce syncytia .
XX
PS Disclosure; Page 44-45; 57pp; French.
XX
CC The present sequence encodes a human endogenous retrovirus envelope
CC protein. The specification describes a method for detecting expression
CC of an envelope protein from a human endogenous retrovirus (HERV), in
CC cells, of a tissue or culture. The method comprises detecting syncytia
CC formation due to the fusogenic properties of the envelope protein.
CC Envelope polypeptides and polynucleotides are used to produce
CC therapeutic or prophylactic compositions, particularly for treatment of
CC cancer, to correct defects in placental development (or other natural
CC formation of other types of syncytia), and to promote adhesion of cells
CC in grafts or cellular repair processes. Expression of sequences
CC antisense to the polynucleotide are used to prevent formation of
CC syncytia.
XX
SQ Sequence 2781 BP; 747 A; 770 C; 556 G; 708 T; 0 other;
Query Match 100.0%; Score 201; DB 22; Length 2781;
Best Local Similarity 100.0%; Pred. No. 6e-55;
Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CCTAGCAGCAGTAGTCTCTCAAAATCGAAGAGCTTTAGACTTGTCAACCGCTGAAGAGG 60
DB 1883 CCTAGCAGCAGTAGTCTCTCAAAATCGAAGAGCTTTAGACTTGTCAACCGCTGAAGAGG 1942
QY 61 GGGAACTGTTATTTTGGGGAAGAATGCTGTATTATTTAATCAATCCGGAATCGT 120
DB 1943 GGGAACTGTTATTTTGGGGAAGAATGCTGTATTATTTAATCAATCCGGAATCGT 2002
QY 121 CACTGAGAAAGTTAAAGAATTCGAGATCGAATACAACTAGACAGAGAGCTTCGAAA 180
DB 2003 CACTGAGAAAGTTAAAGAATTCGAGATCGAATACAACTAGACAGAGAGCTTCGAAA 2062
QY 181 CACTGAGACCTGGGGGCTCTCT 201
DB 2063 CACTGAGACCTGGGGGCTCTCT 2083
RESULT 5
AAD24195
ID AAD24195 standard; cDNA; 2930 BP.
XX
AC AAD24195;
XX
DT 07-MAY-2002 (first entry)
XX
DE Human syncytin cDNA.
XX
XX
XX Human syncytin; preclampsia; gestational trophoblast disorder;
XX choriocarcinoma; hydatiform mole; placental site tumour; abortion;
```

KX		envelope gene; human endogenous defective retrovirus; HERV-W, ss.
XX		
OS	Homo sapiens.	
XX		
FH	Key	Location/Qualifiers
FT	CDS	930..2546 /*tag= a /product= "Syncytin"
XX		
PN	WO200204678-A2.	
XX		
PD	17-JAN-2002.	
XX		
PE	09-JUL-2001; 2001WO-US21719.	
XX		
PR	07-JUL-2000; 2000US-216657P.	
XX		
PA	(GEMV ) GENETICS INST INC.	
XX		
PI	Kelth JC, McCoy JM, Ml S;	
XX		
DR	WPI: 2002-171727/22.	
XX	P-PsDB: AAE14540.	
PT	Identifying a compound for treating a subject with or at risk of	
PP	developing preecclampsia, comprises determining whether the expression	
PT	or activity of syncytin in the cell is modulated in the presence of a	
PT	test compound -	
XX		
PS	Disclosure; page 39-42; 43pp; English.	
XX		
CC	The invention relates to identifying compounds which are modulators	
CC	of syncytin expression. The syncytin modulators are useful in diagnosis	
CC	and treatment of preeclampsia and gestational trophoblast disorders (e.g.	
CC	choriocarcinoma, hydatiform mole, placental site tumour and missed/ incomplete abortion). Syncytin is a human gene derived from the	
CC	envelope gene of human endogenous defective retrovirus, HERV-W. The	
CC	present invention is based partly on the discovery that syncytin	
CC	expression is dramatically reduced in preeclampsia, and is also	
CC	mis-localised to the apical syncytiotrophoblast membrane. The present	
CC	sequence is human syncytin cDNA.	
XX		
SQ	Sequence 2930 BP; 842 A; 800 C; 571 G; 717 T; 0 other:	
	Query Match	100.0%; Score 201; DB 24; Length 2930;
	Best Local Similarity	100.0%; Pred. No. 6.le-55;
	Matches 201; Conservative	0; Mismatches 0; Indels 0; Gaps 0,
OY	1 CCTACAGCAGTAGTCCTTCAAAATCGAAGACTTTAGACTTGCTTACCCTGGAAAGAG	60
Db	2051 CCTACAGCAGTAGTCCTTCAAAATCGAAGACTTTAGACTTGCTTACCCTGGAAAGAG	2110
OY	61 GGGAACCTGTATTATTTTTAGGGAGAATGCTGTATTATGTTTAATCATCCGAATGCT	120
Db	2111 GGGAACCTGTATTATTTTTAGGGAGAATGCTGTATTATGTTTAATCATCCGAATGCT	2170
OY	121 CACTGAGAAAGTTAAAGAAATTCGAGATCGAATACACGTAGAGCAGAGAGCTTCGAAA	180
Db	2171 CACTGAGAAAGTTAAAGAAATTCGAGATCGAATACACGTAGAGCAGAGAGCTTCGAAA	2230
OY	181 CACTGAGACCTTGGGCGCTCT 201	
Db	2231 CACTGAGACCTTGGGCGCTCT 2251	
RESULT 6		
AAK77526		
ID	AAK77526 standard; CDNA; 2946 BP.	
XX	AAK77526;	
XX		
DT	10-AUG-1999 (first entry)	
XX		

DE	Human secreted protein AJ172_2 cDNA.
XX	
KW	Secreted protein; testes; brain; blood; placenta; human; murine; thymus;
KV	bone marrow; treatment; prevention; nutrition; cytokine; immune; vaccine;
KM	cell proliferation; cell differentiation; suppressor; tumour inhibition;
KX	haematopoiesis regulator; activin; inhibin; chemotactic; chemokinetic;
KW	haemostatic; thrombolytic; receptor; ligand; anti-inflammatory; tumour;
KM	cadherin; tumour invasion suppressor; gene therapy; tissue growth; ss.
XX	
OS	Homo sapiens.
XX	
PN	W0926972-A1.
XX	
PD	03-JUN-1999.
XX	
PE	17-NOV-1998; 98WO-US24614.
XX	
PR	20-OCT-1998; 98US-0175928.
PR	21-NOV-1997; 97US-0976110.
BR	18-MAY-1998; 98US-0080478.
XX	
PA	(GENY ) GENETICS INST INC.
XX	
PI	Collins-Racie LA, Evans C, Jacobs K, Lavallie ER;
PI	Mccoy JM, Merberg D, Treacy M,
XX	
DR	WPI: 1999-357813/30.
XX	
DR	P-PSDB: AAY08622.
XX	
PT	New polynucleotides encoding secreted proteins
PS	Claim 13a: Page 100-101; 14zpp; English.
XX	
CC	This invention describes novel human secreted proteins encoded by
CC	polynucleotides isolated from human adult testes, adult brain, adult
CC	blood or adult placenta, or murine adult bone marrow or thymus cDNA
CC	libraries. The products of the invention are predicted to have biological
CC	activities which would make them suitable for treating, preventing or
CC	ameliorating medical conditions in humans and animals, although no
CC	supporting data is given. Suggested activities include nutritional
CC	activity, cytokine and cell proliferation/differentiation activity,
CC	immune stimulating (e.g. as vaccines) or suppressing activity,
CC	haematopoiesis regulating activity, tissue growth activity,
CC	activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic
CC	and thrombolytic activity, receptor/ligand activity, anti-inflammatory
CC	activity, cadherin/tumour invasion suppressor activity, and tumour
CC	inhibition activity. The polynucleotides are also stated to be useful
CC	for gene therapy.
XX	
SQ	Sequence 2946 BP; 858 A; 801 C; 570 G; 717 T; 0 other:
	Query Match            100.0%; Score 201; DB 20; Length 2946;
	Best Local Similarity 100.0%; Pred. No. 6,le-55;
	Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0
OY	1 CCTAGCAGCAGTAGTCTCTCAAAATCGAAGAGCTTTAGACTGTACCCGCTGAAGAAGG 60 
Db	2049 CCTAGCAGCAGTAGTCTCTCAAAATCGAAGAGCTTTAGACTGTACCCGCTGAAGAAGG 2108 
OY	61 GGGAACTGTTTTATTITTTAGGGAGAAAGTGCTTTATTTATGTAATCAATCCGGAATGCT 120 
Db	2109 GGGAACTGTTTTATTITTTAGGGAGAAAGTGCTTTATTTATGTAATCAATCCGGAATGCT 2168 
OY	121 CACTGAGAAAGTTAAACAATAATTCGAGATGGAATACAAAGTACAGAGAGAGACTTCGAAA 180 
Db	2169 CACTGAGAAAGTTAAACAATAATTCGAGATGGAATACAAAGTACAGAGAGAGACTTCGAAA 2228 
OY	181 CACTGAGACCCTGGGGCCTCCT 201 
Db	2229 CACTGAGACCCTGGGGCCTCCT 2249

AA259468  
ID AA259468 standard; cDNA; 2946 BP.  
XX  
AC AA259468;  
XX  
DT 11-APR-2000 (first entry)  
XX  
DE Human secreted protein AJ172\_2 polynucleotide sequence.  
XX  
KW Human; secreted protein; disease diagnosis; pre-eclampsia; cancer;  
KW placental pathology; metastasis inhibition; nutritional activity;  
KW immune stimulator; haematopoiesis regulator; tissue growth;  
KW tumour inhibitor; anti-inflammatory; clone AJ172\_2; ATCC\_98115;  
KW gene therapy; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO960020-A1.  
XX  
PD 25-NOV-1999.  
XX  
PF 17-MAY-1999; 99WO-US10915.  
XX  
PR 18-MAY-1998; 98US-0080478.  
PR 20-OCT-1998; 98US-0175928.  
XX  
PA (GEMT ) GENETICS INST INC.  
XX  
PI Jacobs K, McCoy JM, Lavallee ER, Collins-Racle LA, Evans C;  
PI Merberg D, Ml S, Treacy M;  
XX  
DR WPI: 2000-116311/10.  
DR P-PSDB: AA67313.  
XX  
PT New polynucleotides encoding secreted cDNA libraries, used to develop  
PT products for the diagnosis and treatment of neoplastic disease  
XX  
PS Claim 14; Page 107-108; 149pp; English.  
XX  
CC This is the human secreted protein AJ172\_2 nucleotide sequence, obtained  
CC from a human adult testes cDNA library. The invention relates to secreted  
CC human and murine proteins. The polynucleotides and proteins are predicted  
CC to have biological activities which would make them suitable for  
CC treating, preventing or ameliorating medical conditions in humans and  
CC animals. Detection of the levels of the proteins can be used for the  
CC diagnosis of e.g. pre-eclampsia, placental pathology or cancer. Agents  
CC which modulate the expression or function of the proteins may be used for  
CC treating a neoplastic disease and inhibiting metastasis. Other suggested  
CC activities include nutritional activity (e.g. in feeds), cytokine and  
CC cell proliferation/differentiation activity, immune stimulating  
CC (e.g. as vaccines) or suppressing activity, haematopoiesis regulating  
CC activity, tissue growth activity, activin/inhibin activity,  
CC chemotactic/chemokinetic activity, haemostatic and thrombolytic activity,  
CC receptor/ligand activity, anti-inflammatory activity, cadherin/tumour  
CC invasion suppressor activity, and tumour inhibition activity. The  
CC polynucleotide sequences are also stated to be useful for gene therapy.  
XX  
SQ Sequence 2946 BP; 858 A; 801 C; 570 G; 717 T; 0 other;  
XX  
Query Match 100.0%; Score 201; DB 21; Length 2946;  
Best Local Similarity 100.0%; Pred. No. 6.1e-55;  
Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 2169 CACTGAGAAAGTTAAAGAAATTCGAGATCGAATACACGTAGACGAGAGCTTCGAAA 2228  
QY 181 CACTGAGACCCCTGGGCGCTCCT 201  
DB 2229 CACTGAGACCCCTGGGCGCTCCT 2249  
XX  
DE Human retroviral sequence HERV-7q.  
XX  
KW Autoimmune disease; HERV-7q; chromosome 7q; immunotherapy;  
KW multiple sclerosis; ds.  
XX  
OS Human retrovirus.  
XX  
PN WO967395-A1.  
XX  
PD 29-DEC-1999.  
XX  
PF 23-JUN-1999; 99WO-FR01513.  
XX  
PR 23-JUN-1998; 98FR-0007920.  
XX  
PA (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.  
XX  
PI Alliel PM, Perin J, Rieger F;  
XX  
DR WPI: 2000-160587/14.  
XX  
PT New nucleic acid sequences of human endogenous retrovirus, HERV-7q,  
PT used for diagnosis, treatment and prevention of autoimmune and  
PT neurological diseases  
XX  
PS Claim 3; Fig 1; 225pp; French.  
XX  
CC The present invention relates to new nucleic acid sequences of human  
CC endogenous retrovirus, HERV-7q, which is located on chromosome 7q.  
CC Regulatory elements associated with HERV-7q may alter expression of other  
CC genes (even remote genes) on the same chromosome, inducing immunological  
CC and/or neurological changes (which may be pathological or protective/  
CC curative). HERV-7q peptides can be used to improve efficiency of the  
CC immune response, e.g. in immunotherapy. HERV-7q peptides and their coding  
CC sequences can be used in immunogenic or vaccinating compositions, for  
CC protection against autoimmune diseases, particularly multiple sclerosis.  
CC The peptides may also be used (by sequence comparison) to detect/identify  
CC endogenous retroviruses that are abnormally expressed in cancer,  
CC neuropathologies or other autoimmune diseases. The present sequence was  
CC used to illustrate the invention.  
XX  
SQ Sequence 10499 BP; 3048 A; 2676 C; 2280 G; 2495 T; 0 other;  
XX  
Query Match 100.0%; Score 201; DB 21; Length 10499;  
Best Local Similarity 100.0%; Pred. No. 9.7e-55;  
Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 181 CACTGACCCCTGGGGCTCCT 201  
|||||  
Db 9180 CACTGACCCCTGGGGCTCCT 9200

RESULT 9  
ABL61744  
ID ABL61744 standard; DNA; 56093 BP.  
XX

AC ABL61744;

DT 15-MAY-2002 (first entry)

DE Colon adenocarcinoma related gene sequence SEQ ID NO:81.

XX Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;  
KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;  
KW cytoskeletal; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;  
gene; ds.  
XX

OS Homo sapiens.

PN W0200194629-A2.

PD 13-DEC-2001.

PE 30-MAY-2001; 2001MO-US10838.

PR 05-JUN-2000; 2000US-209473P.

PR 05-JUN-2000; 2000US-209531P.

PR 18-SEP-2000; 2000US-231333P.

PR 18-SEP-2000; 2000US-233617P.

PR 20-SEP-2000; 2000US-234099P.

PR 20-SEP-2000; 2000US-234034P.

PR 20-SEP-2000; 2000US-234052P.

PR 22-SEP-2000; 2000US-234509P.

PR 22-SEP-2000; 2000US-234567P.

PR 25-SEP-2000; 2000US-234923P.

PR 25-SEP-2000; 2000US-234924P.

PR 25-SEP-2000; 2000US-235077P.

PR 25-SEP-2000; 2000US-235082P.

PR 25-SEP-2000; 2000US-235134P.

PR 25-SEP-2000; 2000US-235280P.

PR 26-SEP-2000; 2000US-235637P.

PR 26-SEP-2000; 2000US-235638P.

PR 27-SEP-2000; 2000US-235711P.

PR 27-SEP-2000; 2000US-235720P.

PR 27-SEP-2000; 2000US-235840P.

PR 27-SEP-2000; 2000US-235863P.

PR 28-SEP-2000; 2000US-236028P.

PR 28-SEP-2000; 2000US-236032P.

PR 28-SEP-2000; 2000US-236033P.

PR 28-SEP-2000; 2000US-236034P.

PR 28-SEP-2000; 2000US-236109P.

PR 28-SEP-2000; 2000US-236111P.

PR 29-SEP-2000; 2000US-236842P.

PR 29-SEP-2000; 2000US-236891P.

PR 02-OCT-2000; 2000US-237172P.

PR 02-OCT-2000; 2000US-237173P.

PR 02-OCT-2000; 2000US-237278P.

PR 02-OCT-2000; 2000US-237294P.

PR 02-OCT-2000; 2000US-237295P.

PR 02-OCT-2000; 2000US-237316P.

PR 03-OCT-2000; 2000US-237425P.

PR 03-OCT-2000; 2000US-237596P.

PR 03-OCT-2000; 2000US-237604P.

PR 03-OCT-2000; 2000US-237606P.

PR 03-OCT-2000; 2000US-237608P.

PR 01-NOV-2000; 2000US-244867P.

PR 01-NOV-2000; 2000US-245084P.

XX (AVAL-) AVALON PHARM.

XX Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;  
PI Soppet DR, Weaver Z;  
XX WPI; 2002-168264/24.

XX Screening for anti-neoplastic agent involves exposing cells to a  
PT chemical agent to be tested for anti-neoplastic activity, and  
PT determining a change in expression of a gene of a signature gene set  
XX

PS Claim 1; SEQ ID 81; 44pp; English.

XX The present invention describes a method (M1) for screening for an  
CC anti-neoplastic agent. The method involves exposing cells to a chemical  
CC agent to be tested for anti-neoplastic activity, determining a change in  
CC expression of at least one gene (I) of a signature gene set, where (I)  
CC comprises a sequence (S) selected from 8447 sequences (given in ABL61664  
CC to ABL70110), or is at least 95% identical to (S), where a change in  
CC expression is indicative of anti-neoplastic activity. (I) has cytoskeletal  
CC activity and can be used in gene therapy. M1 can be used for screening  
CC an anti-neoplastic agent, and can be used for producing a product which  
CC is the data collected with respect to the anti-neoplastic agent as a  
CC result of M1, and the data is sufficient to convey the chemical  
CC structure and/or properties of the agent. M1 can be used in the  
CC treatment of cancer such as colon, breast, stomach, lung, thyroid,  
CC oesophageal, ovarian, kidney, prostate or pancreatic cancer.  
CC adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,  
CC infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine  
CC carcinoma, papillary carcinoma and Wilm's tumour.

XX Sequence 56093 BP; 16164 A; 12346 C; 10702 G; 16881 T; 0 other:

XX Query Match 100.0%; Score 201; DB 24; Length 56093;  
XX Best Local Similarity 100.0%; Pred. No. 1.8e-54;  
XX Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCTAGACAGCTAGTCTCTCAAAATCGAAGAGCTTTAGACTTGTCTAACCGCTGAAGAGG 60

Db 37000 CCTAGACAGCTAGTCTCTCAAAATCGAAGAGCTTTAGACTTGTCTAACCGCTGAAGAGG 37059

QY 61 GGGAACTGTTATTTTAAAGGGAAGATGCTGTATATATCAATCGGAATCGT 120

Db 37060 GGGAACTGTTATTTTAAAGGGAAGATGCTGTATATATCAATCGGAATCGT 37119

QY 121 CACTGAGAACTTAAGAAATTCGATCGAATCAACGTAGACAGAGAGCTTCGAA 180

Db 37120 CACTGAGAACTTAAGAAATTCGATCGAATCAACGTAGACAGAGAGCTTCGAA 37179

QY 181 CACTGACCCCTGGGGCTCCT 201

Db 37180 CACTGACCCCTGGGGCTCCT 37200

RESULT 10

AAAX25665  
ID AAAX25665 standard; CDNA to mRNA; 7582 BP.

AC AAAX25665;

DT 21-MAY-1999 (first entry)

DE Complete human endogenous retrovirus W genome.

XX Clone; human endogenous retrovirus; genome; autoimmune disease;

KW multiple sclerosis; rheumatoid polyarthritis; insulin-dependent diabetes;

OS disseminated lupus erythematosus; pregnancy; chromosomal marker; ss.

Human endogenous retrovirus.

XX WO9902696-A1.

XX 21-JAN-1999.

XX

XX

```

PF 06-JUL-1998; 98WO-FR01442.
XX
PR 07-JUL-1997; 97FR-0008815.
XX
XX (INMR ) BIO MERIEUX.
XX
PI Beseme F, Blond JL, Boulton O, Mallet F, Mandrand B;
XX WPI; 1999-120897/10.
DR
XX
XX New nucleic acid sequences from human endogenous retrovirus-W -
PT expressed exclusively in placenta and useful in diagnosis and
PT therapy of autoimmune disease, and abnormal or failed pregnancy
XX
XX Claim 1: Page 71-74; 106pp; French.
PS
XX This sequence represents the complete sequence of the human endogenous
CC retrovirus (HERV) W genome. The nucleic acids, their fragments or
CC peptides encoded by them are markers of autoimmune disease (e.g. multiple
CC sclerosis, rheumatoid polyarthritis, disseminated lupus erythematosus,
CC insulin- dependent diabetes and related pathologies) and of abnormal or
CC unsuccessful pregnancy and can be used as chromosomal markers for
CC susceptibility to these conditions, or proximity markers of genes
CC associated with this susceptibility.
XX
SQ Sequence 7582 BP; 2156 A; 1877 C; 1537 G; 1796 T; 2 U; 214 other;

Query Match 98.6%; Score 198.2; DB 20; Length 7582;
Best Local Similarity 96.5%; Pred. No. 6.9e-54;
Matches 194; Conservative 7; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCGAGCAGCAGTAGTCTTCAAAATCGAAGAGCTTAGACTTGTACCGCTGAAGAGG 60
Db 6702 CCGAGCAGCAGTAGTCTTCAAAATCGAAGAGCTTAGACTTGTACCGCTGAAGAGG 6761
QY 61 GGGAACTGTTTATTTTAAAGGAGAGATGCTGTTATTTATTCATCCGGAATCGT 120
Db 6762 GGGAACTGTTTATTTTAAAGGAGAGATGCTGTTATTTATTCATCCGGAATCGT 6821
QY 121 CACTGAGAAAGTTAAAGAAATTCGAGATCGAATACAGAGAGAGAGCTTCGAAA 180
Db 6822 CACTGAGAAAGTTAAAGAAATTCGAGATCGAATACAGAGAGAGAGCTTCGAAA 6881
QY 181 CACTGAGACCTGGGGCCTCCT 201
Db 6882 CACTGAGACCTGGGGCCTCCT 6902

RESULT 11
AAAS9215
ID AAAS9215 standard; DNA; 7582 BP.
XX
XX AAAS9215;
AC
XX
XX 07-NOV-2000 (first entry)
DT
XX
XX Human endogenous retrovirus W (HERV-W) sequence.
KW
XX
XX Autoimmune disease; retrovirus; human endogenous retrovirus W; HERV-W;
KW gag gene; pregnancy; multiple sclerosis; T cell proliferation; ss.
XX
XX Human endogenous retrovirus.
OS
XX
XX
XX Key Location/Qualifiers
FH 1..120
FT /*tag- a
FT /*note- "R of 5' LTR"
FT LTR 121..575
FT /*tag- b
FT /*note- "U5 of 5' LTR"
FT primer_bind 579..596
FT /*tag- c
FT CDS 5581..7164

```

```

FT /*tag- d
FT /*note- "ORF1 env538"
FT CDS 7039..7194
FT /*tag- e
FT /*note- "ORF2 52 AA"
FT CDS 7112..7255
FT /*tag- f
FT /*note- "ORF3 48 AA"
FT misc_feature 7244..7254
FT /*tag- g
FT /*note- "polypurine tract"
FT LTR 7256..7582
FT /*tag- h
FT /*note- "U3-R of 3' LTR"
FT polyA_signal 7563..7569
FT /*tag- i

XX
XX WO200043521-A2.
XX
XX 27-JUL-2000.
XX
XX 21-JAN-2000; 2000WO-FR00144.
XX
XX 21-JAN-1999; 99FR-0000888.
XX
XX (INMR ) BIO MERIEUX.
XX
XX Paranhos-Baccala G, Mallet F, Voisset C;
XX WPI; 2000-499229/44.
XX
XX New nucleic acid from human endogenous retrovirus, useful e.g. for
PT diagnosis of autoimmune disease and complications of pregnancy,
PT contains at least part of the gag gene
XX
XX Disclosure; Page 49-52; 53pp; French.
XX
XX The present sequence represents an endogenous retrovirus, which is
CC associated with an autoimmune disease, and is integrated into the human
CC genome. The retrovirus is human endogenous retrovirus W (HERV-W). The
CC HERV-W retrovirus is associated with autoimmune disease, failure of
CC pregnancy or disorders of pregnancy. HERV-W nucleic acid fragments, or
CC proteins derived from it, are useful for diagnosis of autoimmune
CC disease (specifically multiple sclerosis) and for monitoring pregnancy.
CC The nucleic acid fragments may also be used for in situ labelling of
CC isolated chromosomes, while the transcription product can be used to
CC study or monitor T cell proliferation in vitro.
XX
XX
SQ Sequence 7582 BP; 2156 A; 1876 C; 1538 G; 1796 T; 216 other;

Query Match 98.6%; Score 198.2; DB 21; Length 7582;
Best Local Similarity 96.5%; Pred. No. 6.9e-54;
Matches 194; Conservative 7; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCGAGCAGCAGTAGTCTTCAAAATCGAAGAGCTTAGACTTGTACCGCTGAAGAGG 60
Db 6702 CCGAGCAGCAGTAGTCTTCAAAATCGAAGAGCTTAGACTTGTACCGCTGAAGAGG 6761
QY 61 GGGAACTGTTTATTTTAAAGGAGAGATGCTGTTATTTATTCATCCGGAATCGT 120
Db 6762 GGGAACTGTTTATTTTAAAGGAGAGATGCTGTTATTTATTCATCCGGAATCGT 6821
QY 121 CACTGAGAAAGTTAAAGAAATTCGAGATCGAATACAGAGAGAGAGCTTCGAAA 180
Db 6822 CACTGAGAAAGTTAAAGAAATTCGAGATCGAATACAGAGAGAGAGCTTCGAAA 6881
QY 181 CACTGAGACCTGGGGCCTCCT 201
Db 6882 CACTGAGACCTGGGGCCTCCT 6902

RESULT 12
AAAX25660

```

ID AAX25660 standard; cDNA to mRNA; 1136 BP.  
XX AAX25660;  
AC  
XX 21-MAY-1999 (first entry)  
DT  
XX Human endogenous retrovirus W clone cl.C4C5.  
DE  
XX Clone: human endogenous retrovirus; genome; autoimmune disease;  
KW multiple sclerosis; rheumatoid polyarthritis; insulin-dependent diabetes;  
KM disseminated lupus erythematosus; pregnancy; chromosomal marker; ss.  
XX  
OS Human endogenous retrovirus.  
XX WO9902696-A1.  
XX 21-JAN-1999.  
XX  
XX 06-JUL-1998; 98WO-FR01442.  
XX  
XX 07-JUL-1997; 97FR-0008815.  
XX  
XX (INMR) BIO MERIEUX.  
XX Beseme F, Blond JL, Bouton O, Mallet F, Mandrand B;  
XX WPI: 1999-120897/10.  
XX  
XX New nucleic acid sequences from human endogenous retrovirus-W -  
PT expressed exclusively in placenta and useful in diagnosis and  
PT therapy of autoimmune disease, and abnormal or failed pregnancy  
XX  
XX Claim 1; Page 59-60; 106pp; French.  
XX  
XX This sequence represents clone cl.C4C5 of the human endogenous retrovirus  
CC (HERV) W genome. The nucleic acids, their fragments or peptides encoded  
CC by them are markers of autoimmune disease (e.g. multiple sclerosis,  
CC rheumatoid polyarthritis, disseminated lupus erythematosus, insulin-  
CC dependent diabetes and related pathologies) and of abnormal or  
CC unsuccessful pregnancy and can be used as chromosomal markers for  
CC susceptibility to these conditions, or proximity markers of genes  
CC associated with this susceptibility.  
XX  
XX Sequence 1136 BP; 336 A; 289 C; 241 G; 270 T; 0 other;  
SO  
Query Match 98.4%; Score 197.8; DB 20; Length 1136;  
Best Local Similarity 99.0%; Pred. No. 4.7e-54;  
Matches 199; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
OY 1 CCTAGCAGCAGTAGTCTTCAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAGAAG 60  
DB 236 CCTAGCAGCAGTAGTCTTCAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAGAAG 295  
OY 61 GGAACCTGTTATTTTAAAGGGAAGATGCTGTTATTTATGTTATCAATCCGAATCGT 120  
DB 296 GGAACCTGTTATTTTAAAGGGAAGATGCTGTTATTTATGTTATCAATCCGAATCGT 355  
OY 121 CACTGAGAAAGTTAAAGAAATTCGAGATCGAATACACGTAGAGAGAGAGCTTCGAAA 180  
DB 356 CACTGAGAAAGTTAAAGAAATTCGAGATCGAATACACGTAGAGAGAGAGCTTCGAAA 415  
OY 181 CACTGAGACCTGGGGCTCTCT 201  
DB 416 CACTGAGACCTGGGGCTCTCT 436  
RESULT 13  
AAS84210  
ID AAS84210 standard; DNA; 1136 BP.  
XX  
XX AAS84210;  
AC  
XX 07-NOV-2000 (first entry)  
DT

XX 3' pol gene and 3' non coding sequences of HERV-W from human genome.  
DE  
XX Autoimmune disease; retrovirus; human endogenous retrovirus W; HERV-W;  
KW gag gene; pregnancy; multiple sclerosis; T cell proliferation; ss.  
XX  
XX Homo sapiens.  
XX  
XX WO200043521-A2.  
XX 27-JUL-2000.  
XX  
XX 21-JAN-2000; 2000WO-FR00144.  
XX  
XX 21-JAN-1999; 99FR-000888.  
XX  
XX (INMR) BIO MERIEUX.  
XX  
XX Paranhos-Baccala G, Mallet F, Volasset C;  
XX WPI: 2000-499229/44.  
XX  
XX New nucleic acid from human endogenous retrovirus, useful e.g. for  
PT diagnosis of autoimmune disease and complications of pregnancy,  
PT contains at least part of the gag gene  
XX  
XX Disclosure; Page 46; 53pp; French.  
XX  
XX The present sequence represents an endogenous retroviral nucleic acid  
CC fragment, which is associated with an autoimmune disease, and is  
CC integrated into the human genome. The fragment is originally derived  
CC from a novel retrovirus, human endogenous retrovirus W (HERV-W). The  
CC HERV-W retrovirus is associated with autoimmune disease, failure of  
CC pregnancy or disorders of pregnancy. The nucleic acid fragment, or  
CC proteins derived from it, are useful for diagnosis of autoimmune  
CC disease (specifically multiple sclerosis) and for monitoring pregnancy.  
CC The nucleic acid fragments may also be used for in situ labelling of  
CC isolated chromosomes, while the transcription product can be used to  
CC study or monitor T cell proliferation in vitro.  
XX  
XX Sequence 1136 BP; 336 A; 289 C; 241 G; 270 T; 0 other;  
SO  
Query Match 98.4%; Score 197.8; DB 21; Length 1136;  
Best Local Similarity 99.0%; Pred. No. 4.7e-54;  
Matches 199; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
OY 1 CCTAGCAGCAGTAGTCTTCAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAGAAG 60  
DB 236 CCTAGCAGCAGTAGTCTTCAAAATCGAAGAGCTTTAGACTTGCTAACCGCTGAAGAAG 295  
OY 61 GGAACCTGTTATTTTAAAGGGAAGATGCTGTTATTTATGTTATCAATCCGAATCGT 120  
DB 296 GGAACCTGTTATTTTAAAGGGAAGATGCTGTTATTTATGTTATCAATCCGAATCGT 355  
OY 121 CACTGAGAAAGTTAAAGAAATTCGAGATCGAATACACGTAGAGAGAGAGCTTCGAAA 180  
DB 356 CACTGAGAAAGTTAAAGAAATTCGAGATCGAATACACGTAGAGAGAGAGCTTCGAAA 415  
OY 181 CACTGAGACCTGGGGCTCTCT 201  
DB 416 CACTGAGACCTGGGGCTCTCT 436  
RESULT 14  
AAS84210  
ID AAS84210 standard; cDNA; 6394 BP.  
XX  
XX AAS84210;  
AC  
XX 13-FEB-2002 (first entry)  
DT  
XX DNA encoding novel human diagnostic protein #20014.  
DE  
XX

KW Human: chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US08631.  
XX  
PR 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;  
XX  
DR WPI; 2001-639362/73.  
XX  
P-PSDB; ABG20023.  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity  
XX  
PS Claim 1; SEQ ID No 20014; 103pp; English.  
XX  
XX The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SO Sequence 6394 BP; 1840 A; 1597 C; 1384 G; 1571 T; 2 other;  
Query Match 98.4%; Score 197.8; DB 23; Length 6394;  
Best Local Similarity 99.0%; Pred. No. 8.8e-54;  
Matches 199; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
QY 1 CCTAGCAGCAGTAGTCTCTTAATAATCGAAGAGCTTTAGACTTGTAAACCGCTGAAGAAG 60  
DB 1157 CCTAGCAGCAGTAGTCTCTTAATAATCGAAGAGCTTTAGACTTGTAAACCGCTGAAGAAG 1216  
QY 61 GGGAACTGTTTATTTTATAGGGAAGAATGCTGTATTATGTTAATCAATCCGAATCGT 120  
DB 1217 GGGAACTGTTTATTTTATAGGGAAGAATGCTGTATTATGTTAATCAATCCGAATCGT 1276  
QY 121 CACTGAGAAAGTTAAAGAAATTCGAGATCGAATCAACAGTAGAGACAGAGAGCTTCGAAA 180  
DB 1277 CACTGAGAAAGTTAAAGAAATTCGAGATCGAATCAACAGTAGAGACAGAGAGCTTCGAAA 1336  
QY 181 CACTGAGACCTGGGGCCTCCT 201  
DB 1337 CACTGAGACCTGGGGCCTCCT 1357

RESULT 15  
AAX25661  
ID AAX25661 standard; cDNA to mRNA; 2782 BP.  
XX  
XX AAX25661;  
AC  
XX  
DT 21-MAY-1999 (first entry)  
XX  
DE Human endogenous retrovirus W clone cl.PH74.  
KW Clone: human endogenous retrovirus; genome; autoimmune disease;  
KW multiple sclerosis; rheumatoid polyarthritis; insulin-dependent diabetes;  
KW disseminated lupus erythematosus; pregnancy; chromosomal marker; ss.  
XX  
OS Human endogenous retrovirus.  
XX  
PN WO902696-A1.  
XX  
PD 21-JAN-1999.  
XX  
PF 06-JUL-1998; 98WO-FR01442.  
XX  
PR 07-JUL-1997; 97FR-0008815.  
XX  
PA (INMR) BIO MERIEUX.  
XX  
PI Beseme F, Blond JL, Boulton O, Mallet F, Mandrand B;  
XX  
DR WPI; 1999-120897/10.  
XX  
PT New nucleic acid sequences from human endogenous retrovirus-W -  
PT expressed exclusively in placenta and useful in diagnosis and  
PT therapy of autoimmune disease, and abnormal or failed pregnancy  
XX  
PS Claim 1; Page 60-63; 106pp; French.  
XX  
XX This sequence represents clone cl.PH74 of the human endogenous retrovirus  
CC (HERV) W genome. The nucleic acids, their fragments or peptides encoded  
CC by them are markers of autoimmune disease (e.g. multiple sclerosis,  
CC rheumatoid polyarthritis, disseminated lupus erythematosus, insulin-  
CC dependent diabetes and related pathologies) and of abnormal or  
CC unsuccessful pregnancy and can be used as chromosomal markers for  
CC susceptibility to these conditions, or proximity markers of genes  
CC associated with this susceptibility.  
XX  
SO Sequence 2782 BP; 741 A; 767 C; 565 G; 709 T; 0 other;  
Query Match 96.0%; Score 193; DB 20; Length 2782;  
Best Local Similarity 97.5%; Pred. No. 2.3e-52;  
Matches 196; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
QY 1 CCTAGCAGCAGTAGTCTCTTAATAATCGAAGAGCTTTAGACTTGTAAACCGCTGAAGAAG 60  
DB 1884 CCTAGCAGCAGTAGTCTCTTAATAATCGAAGAGCTTTAGACTTGTAAACCGCTGAAGAAG 1943  
QY 61 GGGAACTGTTTATTTTATAGGGAAGAATGCTGTATTATGTTAATCAATCCGAATCGT 120  
DB 1944 GGGAACTGTTTATTTTATAGGGAAGAATGCTGTATTATGTTAATCAATCCGAATCGT 2003  
QY 121 CACTGAGAAAGTTAAAGAAATTCGAGATCGAATCAACAGTAGAGACAGAGAGCTTCGAAA 180  
DB 2004 CACTGAGAAAGTTAAAGAAATTCGAGATCGAATCAACAGTAGAGACAGAGAGCTTCGAAA 2063  
QY 181 CACTGAGACCTGGGGCCTCCT 201  
DB 2064 CACTGAGACCTGGGGCCTCCT 2084

Search completed: May 2, 2003, 14:52:10  
Job time : 185.333 secs





GenCore version 5.1.5  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 2, 2003, 14:19:30 ; Search time 1112 Seconds  
(without alignments)  
2927.422 Million cell updates/sec

Title: US-09-719-554-3\_COPY\_9000\_9200

Perfect score: 201  
Sequence: 1 cctgacagacagtagctcctc.....actgacacctgggacctcct 201

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :

EST: \*  
1: em\_estdb: \*  
2: em\_estdb: \*  
3: em\_estdb: \*  
4: em\_estdb: \*  
5: em\_estdb: \*  
6: em\_estdb: \*  
7: em\_estdb: \*  
8: em\_estdb: \*  
9: em\_estdb: \*  
10: em\_estdb: \*  
11: em\_estdb: \*  
12: em\_estdb: \*  
13: em\_estdb: \*  
14: em\_estdb: \*  
15: em\_estdb: \*  
16: em\_estdb: \*  
17: em\_estdb: \*  
18: em\_estdb: \*  
19: em\_estdb: \*  
20: em\_estdb: \*  
21: em\_estdb: \*  
22: em\_estdb: \*  
23: em\_estdb: \*  
24: em\_estdb: \*  
25: em\_estdb: \*  
26: em\_estdb: \*  
27: em\_estdb: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	201	100.0	322	12	BF990902 RC6-GN007
2	201	100.0	411	12	BG004254 RC6-GN007
3	201	100.0	494	14	BQ365207 RC6-GN007
4	201	100.0	714	12	BE734284 601565487
5	200	99.5	354	14	BQ365139 RC6-GN007
6	198.4	98.7	367	12	BF990707 RC6-GN007

C	7	197.8	98.4	346	14	BQ365143	BQ365143 RC6-GN007
C	8	187	93.0	702	9	AU138097	AU138097 AU138097
C	9	185	92.0	310	14	BQ365206	BQ365206 RC6-GN007
C	10	172	85.6	385	17	AQ062249	AQ062249 CTT-HSP-2
C	11	167	83.1	586	10	BE019603	BE019603 ba84f03.y
C	12	163.2	81.2	572	14	BQ365363	BQ365363 MR2-GN002
C	13	162.8	81.0	302	12	BG004247	BG004247 RC6-GN007
C	14	157.8	78.5	608	17	AQ072469	AQ072469 HS_5403.A
C	15	152.4	75.8	790	12	BG572445	BG572445 602593490
C	16	149.8	74.5	294	14	R01948	R01948 ye85b02.r1
C	17	143.8	71.5	723	9	AU138405	AU138405 AU138405
C	18	135.4	67.4	289	13	B1053219	B1053219 PMO-GN021
C	19	133.8	66.6	316	13	B1053220	B1053220 PMO-GN021
C	20	129.8	64.6	305	12	BF990710	BF990710 RC6-GN007
C	21	127	63.2	427	9	AA776439	AA776439 zj50h06.s
C	22	125	62.2	433	14	N77302	N77302 yv43e03.r1
C	23	122.4	60.9	449	14	N58614	N58614 yv56h11.r1
C	24	106.6	53.0	385	14	T93615	T93615 ye05g09.s1
C	25	102	50.7	666	13	BG911940	BG911940 602809647
C	26	99.4	49.5	665	17	AG066067	AG066067 Pan trogl
C	27	98.8	49.2	323	14	R00744	R00744 ye74h09.s1
C	28	93	46.3	422	17	AQ150723	AQ150723 HS_3203.A
C	29	82	40.8	264	17	AQ151165	AQ151165 HS_2186.A
C	30	79.8	39.7	314	9	AA960921	AA960921 oos1a07.s
C	31	77	38.3	771	13	B1087886	B1087886 602852690
C	32	76.8	38.2	314	17	AQ539232	AQ539232 RPT-11-3
C	33	74.4	37.0	179	14	T95945	T95945 ye42e02.r1
C	34	74.4	37.0	664	17	AG090771	AG090771 Pan trogl
C	35	73.4	36.5	616	10	AM847812	AM847812 IL3-CT021
C	36	71.8	35.7	471	10	AM851794	AM851794 OVO-CT022
C	37	70	34.8	634	10	BB635247	BB635247 B635247
C	38	70	34.8	710	17	B69218	B69218 CTT-HSP-205
C	39	68	33.8	636	17	AG116240	AG116240 Pan trogl
C	40	67.2	33.4	754	17	AG031563	AG031563 Pan trogl
C	41	67	33.3	374	14	BQ346732	BQ346732 IL5-NF007
C	42	66.2	32.9	370	12	BG011064	BG011064 OVI-GN031
C	43	65.8	32.7	414	9	AT508499	AT508499 vb34d08.y
C	44	65.8	32.7	425	17	AQ558380	AQ558380 HS_2068.B
C	45	65.4	32.5	361	13	BI977020	BI977020 486270 MA

## ALIGNMENTS

RESULT 1  
LOCUS BF990902 322 bp mRNA linear EST 23-JAN-2001  
DEFINITION RC6-GN0070-271000-023-H12 GN0070 Homo sapiens CDNA, mRNA sequence.  
ACCESSION BF990902  
VERSION BF990902.1 GI:12397227  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE  
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Britones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bata,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare

M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.

Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
20202663  
Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil

Tel: +55-11-2704922  
Fax: +55-11-2707001

Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=RC6&lt2-RC6-GN0070-271000-023-H12&lt3-2000-10-27&lt4-1)  
Seq primer: puc 18 forward  
High quality sequence start: 22  
High quality sequence stop: 275.

## FEATURES

Location/Qualifiers

1..322

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_id="GN0070"

/dev\_stage="Adult"

/note="Organ: placenta.normal; Vector: puc18; Site\_1: SmaI; Site\_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent Application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

## BASE COUNT

95 a 77 c 72 g 78 t

## ORIGIN

## Query Match

100.0%; Score 201; DB 12; Length 322;

Best Local Similarity 100.0%; Pred. No. 2.2e-51; Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCTAGCAGCAGTACTCCTTCAAAATCGAAGAGCTTTAGACTTGTCAACCGCTGAAGAAG 60

DB 88 CCTAGCAGCAGTACTCCTTCAAAATCGAAGAGCTTTAGACTTGTCAACCGCTGAAGAAG 147

QY 61 GGGAACTGTTATTTTATAGGGAAGAATGCTGTATTATGTATCAATCCGAATCGT 120

DB 148 GGGAACTGTTATTTTATAGGGAAGAATGCTGTATTATGTATCAATCCGAATCGT 207

QY 121 CACTGAGAAAGTTAAAGAAATTCGAGATCGAATCAAGCTGAGCAGAGAGCTTCGAAA 180

DB 208 CACTGAGAAAGTTAAAGAAATTCGAGATCGAATCAAGCTGAGCAGAGAGCTTCGAAA 267

QY 181 CACTGAGCCTGGGGCCTCCT 201

DB 268 CACTGAGCCTGGGGCCTCCT 288

RESULT 2

BG004254/c 411 bp mRNA linear EST 24-JAN-2001

LOCUS BG004254

DEFINITION RC6-GN0070-301100-014-C06 GN0070 Homo sapiens cDNA, mRNA sequence.

ACCESSION BG004254

VERSION BG004254.1 GI:12445242

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 411)

AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H., Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare

M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.

Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

JOURNAL MEDLINE  
CONTACT: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922.

Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=RC6&lt2-RC6-GN0070-301100-014-C06&lt3-2000-11-30&lt4-1)  
Seq primer: puc 18 forward  
High quality sequence start: 21  
High quality sequence stop: 411.

## FEATURES

Location/Qualifiers

1..411

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone\_id="GN0070"

/dev\_stage="Adult"

/note="Organ: placenta.normal; Vector: puc18; Site\_1: SmaI; Site\_2: SmaI; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent Application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

## BASE COUNT

103 a 94 c 96 g 118 t

## ORIGIN

## Query Match

100.0%; Score 201; DB 12; Length 411;

Best Local Similarity 100.0%; Pred. No. 2.3e-51; Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCTAGCAGCAGTACTCCTTCAAAATCGAAGAGCTTTAGACTTGTCAACCGCTGAAGAAG 60

DB 294 CCTAGCAGCAGTACTCCTTCAAAATCGAAGAGCTTTAGACTTGTCAACCGCTGAAGAAG 235

QY 61 GGGAACTGTTATTTTATAGGGAAGAATGCTGTATTATGTATCAATCCGAATCGT 120

DB 234 GGGAACTGTTATTTTATAGGGAAGAATGCTGTATTATGTATCAATCCGAATCGT 175

QY 121 CACTGAGAAAGTTAAAGAAATTCGAGATCGAATCAAGCTGAGCAGAGAGCTTCGAAA 180

DB 174 CACTGAGAAAGTTAAAGAAATTCGAGATCGAATCAAGCTGAGCAGAGAGCTTCGAAA 115

QY 181 CACTGAGCCTGGGGCCTCCT 201

DB 114 CACTGAGCCTGGGGCCTCCT 94

RESULT 3

B0365207/c 494 bp mRNA linear EST 21-MAY-2002

LOCUS B0365207

DEFINITION RC6-GN0070-220800-021-D12 GN0070 Homo sapiens cDNA, mRNA sequence.

ACCESSION B0365207

VERSION B0365207.1 GI:21040719

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 494)

AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H., Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare

M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.

Shotgun sequencing of the human transcriptome with ORF expressed sequence tags

JOURNAL MEDLINE  
CONTACT: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil

Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
(<http://www.ludwig.org.br/scripts/gethtml2.pl?cl=RC6&cl2=RC6-GN0070-220800-021-D12&cl3=2000-08-22&cl4=1>)  
Seq primer: puc 18 forward  
High quality sequence start: 40  
High quality sequence stop: 432.  
Location/Qualifiers

## FEATURES

source

1. .494  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="GN0070"  
/dev\_stage="Adult"  
/note="Organ: placenta\_normal; Vector: puc18; Site\_1: SmaI  
; Site\_2: SmaI; A mini-library was made by cloning  
products derived from ORESTES PCR (U.S. Letters Patent  
Application No. 196,716 - Ludwig Institute for Cancer  
Research) profiles into the puc 18 vector. Reverse  
transcription of tissue mRNA and cDNA amplification were  
performed under low stringency conditions."

## BASE COUNT

125 a 105 c 121 g 143 t

## ORIGIN

Query Match 100.0%; Score 201; DB 14; Length 494;  
Best Local Similarity 100.0%; Pred. No. 2,4e-51;  
Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CCTAGCAGCAGTAGTCCTTCAAAATCGAAGAGCTTACCTGCTACCGCTGAAGAGG 60  
|||||  
DB 400 CCTAGCAGCAGTAGTCCTTCAAAATCGAAGAGCTTACCTGCTACCGCTGAAGAGG 341  
|||||  
OY 61 GGGAACTGTTATTTTATTTAGGGGAGAAATGCTGTTATTTATGTTAATCAATCCGAATCGT 120  
|||||  
DB 340 GGGAACTGTTATTTTATTTAGGGGAGAAATGCTGTTATTTATGTTAATCAATCCGAATCGT 281  
|||||  
OY 121 CACTGAGAAAGTTAAAGAAATTCGAGATCGAATACAGCTAGAGAGAGAGACTTCGAAA 180  
|||||  
DB 280 CACTGAGAAAGTTAAAGAAATTCGAGATCGAATACAGCTAGAGAGAGAGACTTCGAAA 221  
|||||  
OY 181 CACTGGACCTGGGGCCTCTCT 201  
|||||  
DB 220 CACTGGACCTGGGGCCTCTCT 200  
|||||

## RESULT 4

BE734284

LOCUS 601565487F1 NIH\_MGC\_21 Homo sapiens cDNA clone IMAGE:3840572 5',  
DEFINITION mRNA sequence.  
ACCESSION BE734284  
VERSION BE734284.1 GI:10148276  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 714)  
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.  
TITLE "National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgabbs-remail.nih.gov  
Tissue Procurement: ATCC  
cDNA Library Preparation: Ling Hong/Rubin Laboratory  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMN)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLMN at: [image.llnl.gov](http://image.llnl.gov)  
Plate: L1CM530 row: m column: 21  
High quality sequence stop: 712.

## FEATURES

source

Location/Qualifiers

1. .714  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="IMAGE:3840572"  
/clone\_lib="NIH\_MGC\_21"  
/tissue\_type="choriocarcinoma"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: placenta; Vector: pOTB7; Site\_1: XhoI;  
Site\_2: EcoRI; cDNA made by oligo-dT priming.  
directionally cloned into EcoRI/XhoI sites using the  
following 5' adaptor: GGCACGAG(G). Size-selected >500bp  
for average insert size 1.8kb. library constructed by  
Ling Hong in the laboratory of Gerald M. Rubin (University  
of California, Berkeley) using ZAP-cDNA synthesis kit  
(Stratagene) and Superscript II RT (Life Technologies)."

## BASE COUNT

194 a 183 c 137 g 200 t

## ORIGIN

Query Match 100.0%; Score 201; DB 12; Length 714;  
Best Local Similarity 100.0%; Pred. No. 2,6e-51;  
Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CCTAGCAGCAGTAGTCCTTCAAAATCGAAGAGCTTACCTGCTACCGCTGAAGAGG 60  
|||||  
DB 397 CCTAGCAGCAGTAGTCCTTCAAAATCGAAGAGCTTACCTGCTACCGCTGAAGAGG 456  
|||||  
OY 61 GGGAACTGTTATTTTATTTAGGGGAGAAATGCTGTTATTTATGTTAATCAATCCGAATCGT 120  
|||||  
DB 457 GGGAACTGTTATTTTATTTAGGGGAGAAATGCTGTTATTTATGTTAATCAATCCGAATCGT 516  
|||||  
OY 121 CACTGAGAAAGTTAAAGAAATTCGAGATCGAATACAGCTAGAGAGAGAGACTTCGAAA 180  
|||||  
DB 517 CACTGAGAAAGTTAAAGAAATTCGAGATCGAATACAGCTAGAGAGAGAGACTTCGAAA 576  
|||||  
OY 181 CACTGGACCTGGGGCCTCTCT 201  
|||||  
DB 577 CACTGGACCTGGGGCCTCTCT 597  
|||||

## RESULT 5

BQ365139

LOCUS BQ365139 354 bp mRNA linear EST 21-MAY-2002  
DEFINITION RC6-GN0070-170800-011-D03 GN0070 Homo sapiens cDNA, mRNA sequence.  
ACCESSION BQ365139  
VERSION BQ365139.1 GI:21040651  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 354)  
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,  
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,  
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bais,G.S., Simpson,D.H.,  
Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare  
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and  
Simpson,A.J.  
TITLE Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
MEDLINE 20202653  
COMMENT Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
(<http://www.ludwig.org.br/scripts/gethtml2.pl?cl=RC6&cl2=RC6-GN0070->

170800-011-D03&f3=2000-08-17&f4=1)

Seq primer: puc 18 forward  
High quality sequence start: 16  
High quality sequence stop: 354.

## FEATURES

Location/Qualifiers

1..354

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_id="GN0070"  
/dev\_stage="Adult"

/note="Organ: placenta\_normal; Vector: puc18; Site\_1: Sma1  
; Site\_2: Sma1; A mini-library was made by cloning  
products derived from ORESTES PCR (U.S. Letters Patent  
Application No. 196,716 - Ludwig Institute for Cancer  
Research) profiles into the puc 18 vector. Reverse  
transcription of tissue mRNA and cDNA amplification were  
performed under low stringency conditions."

BASE COUNT 102 a 86 c 79 g 86 t 1 others

## ORIGIN

Query Match 99.5%; Score 200; DB 14; Length 354;

Best Local Similarity 99.5%; Pred. No. 4,6e-51;

Matches 200; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 CCTAGCAGCAGTACTCTTCAAAATCGAAGAGCTTTAGACTTGTACCGCTGAAGAG 60

DB 120 CCTAGCAGCAGTACTCTTCAAAATCGAAGAGCTTTAGACTTGTACCGCTGAAGAG 179

QY 61 GGAACCTGTTATTTTATAGGGAAGAAATGCTGTTATTTATTAATCAATCCGAATCGT 120

DB 180 GGAACCTGTTATTTTATAGGGAAGAAATGCTGTTATTTATTAATCAATCCGAATCGT 239

QY 121 CACTGAGAAAGTTAAAGAAATTCGAGATCGAATCAACGTAAGACAGAGAGCTTCGAA 180

DB 240 CACTGAGAAAGTTAAAGAAATTCGAGATCGAATCAACGTAAGACAGAGAGCTTCGAA 299

QY 181 CACTGAGACCTTGGGGCCTCT 201

DB 300 CACTGAGACCTTGGGGCCTCT 320

RESULT 6 367 bp mRNA linear EST 23-JAN-2001

BF990707

LOCUS RC6-GN0070-271000-013-D01 GN0070 Homo sapiens cDNA, mRNA sequence.

DEFINITION BF990707.1 GI:12397032

ACCESSION EST.

VERSION human.

KEYWORDS Homo sapiens

SOURCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

ORGANISM Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 367)

AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,

Nagal,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,

Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,

Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare

,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and

Simpson,A.J.

Shotgun sequencing of the human transcriptome with ORF expressed

sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

20202663

CONTACT: Simpson A.J.G.

Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,

Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome

Project. This entry can be seen in the following URL

(http://www.ludwig.org.br/scripts/gethtml2.pl?l1=RC6&f2=RC6-GN0070-271000-013-D01&f3=2000-10-27&f4=1)

Seq primer: puc 18 forward  
High quality sequence start: 17  
High quality sequence stop: 359.

## FEATURES

Location/Qualifiers

1..367

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_id="GN0070"  
/dev\_stage="Adult"

/note="Organ: placenta\_normal; Vector: puc18; Site\_1: Sma1  
; Site\_2: Sma1; A mini-library was made by cloning  
products derived from ORESTES PCR (U.S. Letters Patent  
Application No. 196,716 - Ludwig Institute for Cancer  
Research) profiles into the puc 18 vector. Reverse  
transcription of tissue mRNA and cDNA amplification were  
performed under low stringency conditions."

BASE COUNT 104 a 88 c 85 g 89 t 1 others

## ORIGIN

Query Match 98.7%; Score 198.4; DB 12; Length 367;

Best Local Similarity 99.0%; Pred. No. 1,4e-50;

Matches 199; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CCTAGCAGCAGTACTCTTCAAAATCGAAGAGCTTTAGACTTGTACCGCTGAAGAG 60

DB 133 CCTAGCAGCAGTACTCTTCAAAATCGAAGAGCTTTAGACTTGTACCGCTGAAGAG 192

QY 61 GGAACCTGTTATTTTATAGGGAAGAAATGCTGTTATTTATTAATCAATCCGAATCGT 120

DB 193 GGAACCTGTTATTTTATAGGGAAGAAATGCTGTTATTTATTAATCAATCCGAATCGT 252

QY 121 CACTGAGAAAGTTAAAGAAATTCGAGATCGAATCAACGTAAGACAGAGAGCTTCGAA 180

DB 253 CACTGAGAAAGTTAAAGAAATTCGAGATCGAATCAACGTAAGACAGAGAGCTTCGAA 312

QY 181 CACTGAGACCTTGGGGCCTCT 201

DB 313 CACTGAGACCTTGGGGCCTCT 333

RESULT 7 346 bp mRNA linear EST 21-MAY-2002

BQ365143/c

LOCUS BQ365143 RC6-GN0070-170800-011-F12 GN0070 Homo sapiens cDNA, mRNA sequence.

DEFINITION BQ365143

ACCESSION BQ365143

VERSION BQ365143.1 GI:21040655

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 346)

AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,

Nagal,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,

Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,

Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare

,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and

Simpson,A.J.

Shotgun sequencing of the human transcriptome with ORF expressed

sequence tags

Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)

20202663

CONTACT: Simpson A.J.G.

Laboratory of Cancer Genetics

Ludwig Institute for Cancer Research

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,

Brazil

Tel: +55-11-2704922

Fax: +55-11-2707001

Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome

Project. This entry can be seen in the following URL:  
(http://www.ludwig.org.br/scripts/gethtml2.pl?l=RC6&t2=RC6-GN0070-170800-011-F1&t3=2000-08-17&t4=1)

Seq primer: puc 18 forward  
High quality sequence start: 18  
High quality sequence stop: 346.  
Location/Qualifiers

## FEATURES

source 1..346

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="GN0070"  
/dev\_stage="Adult"  
/note="Organ: Placenta normal; Vector: puc18; Site: 1: SmaI  
; Site 2: SmaI. A mini-library was made by cloning  
products derived from OREBES PCR (U.S. Letters Patent  
Application No. 196,716 - Ludwig Institute for Cancer  
Research) profiles into the pUC 18 vector. Reverse  
transcription of tissue mRNA and cDNA amplification were  
performed under low stringency conditions."

## BASE COUNT

84 a 84 c 79 g 99 t

## ORIGIN

Query Match 98.4%; Score 197.8; DB 14; Length 346;  
Best Local Similarity 99.0%; Pred. No. 2.2e-50;  
Matches 199; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 CCTAGACGACGATGCTCTTCAAAATCGAAGAGCTTACCTTACCGCTAAAGAGG 60  
|||||  
DB 229 CCTAGACGACGATGCTCTTCAAAATCGAAGAGCTTACCTTACCGCTAAAGAGG 170  
|||||  
QY 61 GGGACCTGTTTATTTTGGGAGAGATGCTGTTTATTTATTAATCAATCCGGAATCGT 120  
|||||  
DB 169 GGGAGCTGTTTATTTTGGGAGAGATGCTGTTTATTTATTAATCAATCCGGAATCGT 110  
|||||  
QY 121 CACTGAGAAAGTTAAAGAAATTCGAGATCGAATATCAACGTAAGCAGAGAGCTTCGAAA 180  
|||||  
DB 109 CACTGAGAAAGTTAAAGAAATTCGAGATCGAATATCAACGTAAGCAGAGAGCTTCGAAA 50  
|||||  
QY 181 CACTGACCTGGGGCTCTCT 201  
|||||  
DB 49 CACTGACCTGGGGCTCTCT 29

## RESULT 8

## LOCUS

AUI38097 702 bp mRNA linear EST 02-AUG-2002

## DEFINITION

AUI38097 PLACE1 Homo sapiens cDNA clone PLACE1007839 5', mRNA  
sequence.

## ACCESSION

AUI38097  
AUI38097.1 GI:10999618

## VERSION

## KEYWORDS

## SOURCE

human.

## ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

## REFERENCE

1 (bases 1 to 702)  
Ota, T., Nishikawa, T., Suzuki, Y., Ishii, S., Saito, K., Kawai, Y.,  
Yamamoto, J., Wakamatsu, A., Nakamura, Y., Nagai, T., Sugano, S. and  
Isogai, T.

## AUTHORS

TITLE  
JOURNAL  
COMMENT  
HRI human cDNA project  
Unpublished (2000)  
Contact: Takao Isogai  
Genomics Laboratory  
Helix Research Institute  
1532-3 Yana, Kisarazu, Chiba 292-0812, Japan  
Tel: 81-438-52-3975  
Fax: 81-438-52-3986  
Email: genomics@hri.co.jp

HRI human cDNA project; 5'- & 3'-end one pass sequencing; Helix  
Research Institute; cDNA library construction; Department of  
Virology, Institute of Medical Science, University of Tokyo, and  
Helix Research Institute.  
Location/Qualifiers

## FEATURES

## source

1..702

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="PLACE1007839"  
/clone\_lib="PLACE1"  
/tissue\_type="Placenta"  
/note="Vector: pME18SFL3"

## BASE COUNT

206 a 173 c 131 g 189 t 3 others

## ORIGIN

Query Match 93.0%; Score 187; DB 9; Length 702;  
Best Local Similarity 98.0%; Pred. No. 5.3e-47;  
Matches 188; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

QY 1 CCTAGACGACGATGCTCTTCAAAATCGAAGAGCTTACCTTACCGCTAAAGAGG 60  
|||||  
DB 490 CCTAGACGACGATGCTCTTCAAAATCGAAGAGCTTACCTTACCGCTAAAGAGG 549  
|||||  
QY 61 GGGACCTGTTTATTTTGGGAGAGATGCTGTTTATTTATTAATCAATCCGGAATCGT 120  
|||||  
DB 550 GGGACCTGTTTATTTTGGGAGAGATGCTGTTTATTTATTAATCAATCCGGAATCGT 609  
|||||  
QY 121 CACTGAGAAAGTTAAAGAAATTCGAGATCGAATATCAACGTAAGCAGAGAGCTTCGAA 179  
|||||  
DB 610 CACTGAGAAAGTTAAAGAAATTCGAGATCGAATATCAACGTAAGCAGAGAGCTTCGAA 669  
|||||  
QY 180 ACACTGGACCTGGGGCTCTCT 201  
|||||  
DB 670 ACACTGGACCTGGGGCTCTCT 691

## RESULT 9

## LOCUS

BQ365206/c 310 bp mRNA linear EST 21-MAY-2002

## DEFINITION

BQ365206 RC6-GN0070-220800-021-D11 GN0070 Homo sapiens cDNA, mRNA sequence.

## ACCESSION

BQ365206.1 GI:21040718

## VERSION

## KEYWORDS

EST.

## SOURCE

human.

## ORGANISM

Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

## REFERENCE

1 (bases 1 to 310)  
Dias Neto, E., Garcia Correa, R., Verjovsky-Almeida, S., Briones, M.R.,  
Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F.,  
Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H.,  
Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare,  
M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and  
Simpson, A.J.

## AUTHORS

TITLE  
JOURNAL  
MEDLINE  
COMMENT  
Shotgun sequencing of the human transcriptome with ORF expressed  
sequence tags  
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
20202663  
Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,  
Brazil  
Tel: +55-11-2704922  
Fax: +55-11-2707001  
Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome  
Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?l=RC6&t2=RC6-GN0070-  
220800-021-D11&t3=2000-08-22&t4=1)  
Seq primer: puc 18 forward  
High quality sequence start: 19  
High quality sequence stop: 310.  
Location/Qualifiers

## FEATURES

source 1..310

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="GN0070"

```

/dev_stage="Adult"
/note="Organ: Placenta.normal; Vector: puc18; Site_1: SmaI
; Site_2: SmaI; A mini-library was made by cloning
products derived from ORESSES PCR (U.S. Letters Patent
Application No. 196,716 - Ludwig Institute for Cancer
Research) profiles into the pUC18 vector. Reverse
transcription of tissue mRNA and cDNA amplification were
performed under low stringency conditions."
BASE COUNT      74 a      73 c      68 g      95 t
ORIGIN

Query Match      92.0%; Score 185; DB 14; Length 310;
Best Local Similarity 97.4%; Pred. No. 1.9e-46;
Matches 188; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 CCTACAGACGATGCTTCCTCAAAATCGAAGACCTTAGCTGCTAACCGCTGAAAGAG 60
    |||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 193 CCTACAGACGATGCTTCCTCAAAATCGAAGACCTTAGCTGCTAACCGCTGAAAGAG 134
    |||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 61 GGGAACTGTATTATTTTGGGGAAGAAATGCTTATTATTTAATCAATCCGGAATCGT 120
    |||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 133 GGGAACTGTATTATTTTGGGGAAGAAATGCTTATTATTTAATCAATCCGGAATCGT 74
    |||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 121 CACTGAGAAAGTTAAAGAAATTCGAGATCGAATACACGTAGACAGAGAGCTTCGAAA 180
    |||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 73 CACTGAGAAAGTTAAAGAAATTCGAGATCGAATACACGTAGACAGAGAGAGCTTCGAAA 14
    |||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 181 CACTGAGACCCCTGG 193
    |||||||||||||||
DB 13 CAGTGGCCCTAGG 1

RESULT 10
A0062249/c 385 bp DNA linear GSS 31-JUL-1998
LOCUS CIT-HSP-2347N16.TR CIT-HSP Homo sapiens genomic clone 2347N16, DNA
DEFINITION sequence.
ACCESSION A0062249
VERSION A0062249.1 GI:3364161
KEYWORDS GSS.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 385)
Adams, M.D., Rounsley, S.D., Zhao, S., Field, C.E., Bass, S., Linher, K.,
Golden, K., Berry, K., Granger, D., Suh, E., Wible, C., Shizuya, H.,
Simon, M., and Venter, J.C.
Use of a random BAC End Sequence Database for Sequence-Ready Map
Building (1998)
Unpublished (1998)
Other GSSs: CIT-HSP-2347N16.TF
Contact: Mark Adams
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: mdamas@tigr.org
Clones are available from Research Genetics (Info@resgen.com). BAC
end search page:
http://www.tigr.org/tdb/hungen/bac_end_search/bac_end_search.html.
Seq primer: M13 Reverse
Class: BAC ends.

FEATURES
Source location/Qualifiers
1..385
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="2347N16"
/clone_lib="CIT-HSP"
/sex="Male"
/note="Vector: pBelobAC11; Site_1: HindIII; Site_2:

```

```

BASE COUNT      109 a      80 c      91 g      105 t
ORIGIN HindIII"

Query Match      85.6%; Score 172; DB 17; Length 385;
Best Local Similarity 93.6%; Pred. No. 2e-42;
Matches 191; Conservative 0; Mismatches 10; Indels 3; Gaps 1;

QY 1 CCTACAGACGATGCTTCCTCAAAATCGAAGACCTTAGACT---GCTAACCGTGAAG 57
    |||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 334 CCTACAGACGATGCTTCCTCAAAATCGAAGACCTTAGACTTAGCTTAATGCCAAG 275
    |||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 58 AGGGGAACCTGTATTATTTTGGGGAAGAAATGCTTATTATTTAATCAATCCGGAAT 117
    |||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 274 AGGGGAACCTGTATTATTTTGGGGAAGAAATGCTTATTATTTAATCAATCCGGAAT 215
    |||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 118 CGTCACTGAGAAAGTTAAAGAAATTCGAGATCGAATACACGTAGACAGAGAGCTTCG 177
    |||||||||||||||||||||||||||||||||||||||||||||||||||||||
DB 214 CGTCACTGAGACAGTTAAAGAAATTCGAGATTCATACACATAGACAGAGAGCTTCA 155
    |||||||||||||||||||||||||||||||||||||||||||||||||||||||
QY 178 AACACGTGACCCCTGGGGCTCCT 201
    |||||||||||||||||||
DB 154 AACACGTGACCCCTGGGGCTCCT 131
    |||||||||||||||||||

RESULT 11
BE019603 586 bp mRNA linear EST 06-JUN-2000
LOCUS ba84f03.y1 NIH_MGC_21 Homo sapiens cDNA clone IMAGE:2907101 5'
DEFINITION similar to TR:095244 095244 ENVELOPE PROTEIN; contains Alu
repetitive element; mRNA sequence.
ACCESSION BE019603
VERSION BE019603.1 GI:8279682
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 586)
NIH-MGC http://mgi.nci.nih.gov/.
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cga@bbs.femail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNLN at:
image.llnl.gov/image/html/resources.shtml
Seq primer: -40RP from Gibco
High quality sequence stop: 519.

FEATURES
Source location/Qualifiers
1..586
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:2907101"
/clone_lib="NIH_MGC_21"
/tissue_type="choriocarcinoma"
/lab_host="DH10B (phage-resistant)"
/note="Organ: Placenta; Vector: pOTB7; Site_1: XhoI;
Site_2: EcoRI; cDNA made by oligo-dT priming.
directionally cloned into EcoRI/XhoI sites using the
following 5' adaptor: GGCAGAG(G). Size-selected >500bp
for average insert size 1.8kb. Library constructed by
Ling Hong in the laboratory of Gerald M. Rubin (University
of California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
```

```

BASE COUNT      168 a      141 c      113 g      164 t
ORIGIN

Query Match      83.1%; Score 167; DB 10; Length 586;

```

Qy 1 CCTAGACAGTAGTCCTTCAAAATCGAAGAGCTTTAG-ACCTGCTAACCGCTGAAGAG 59

— — — — —

[illegible]

Db 195 TCCTGAGAACTTAAAGAAATTCGAGATCGAATACAAGGTAGAGCAGAGGAGCTTCGA 25

QY 180 A C A C T G G A C C C T G G G G C C T C C T 201

Db 255 ACACGGGACCCGGGGCCTCCT 276

BG004247  
LOCUS  
PC004247  
303 bp  
MWNA 14:0000  
ECM 24-TAN-2001

ACCESSION	BC004247
DEFINITION	RC6-GN0070-301100-014-F12 GN0070 Homo sapiens cDNA, mRNA sequence.

KEYWORDS EST, BG00424/.1 GI:1244322/

ORGANISM	Homo sapiens
DOORCE	Human

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

**AUTHORS** Dias Neto, E., Garcia Correa, R., Verjovski-Almeida, S., Briones, M.

Goldman, G.H., Carvalho, A.F., Matsukuma, A., Baia, G.S., Simpson, D., Brunetto, A., and Todorova, D. 2005. The effects of the 2004 Indian Ocean tsunami on the coastal environment of the Andaman Islands, India. *Journal of Coastal Research* 21: 103-114.

M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simson, A.T.

### LINE sequencing of the human transcriptome with ORF sequence tags

MEDLINE 20202663

Laboratory of Cancer Genetics

Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP  
Brasil

Tel: +55-11-2704922  
Fax: +55-11-3707001

This sequence was derived from the FAPESP/LTCR Human Cancer Genom  
Email: [asimpson@luadwig.org.br](mailto:asimpson@luadwig.org.br)

Project. This entry can be seen in the following URL  
(<http://www.ludwig.org.br/scripts/qethml2.pl?l=RC6&t2=RC6-GN00>

Seq primer: puc 18 forward

High quality sequence stop: 301.

source 1. .302

```
/db_xref="taxon:9606"  
/clone_1b="CN0070"
```

```

/dev_stage="Adult"
/notes="Organ: placenta normal: Vector: nuc18: site 1: S

```

site\_2: smai, a mini-libdialy was made by cloning products derived from ORESTES PCR (U.S. Letters Patent

Research) profiles into the pUC 18 vector. Reverse

performed under low stringency conditions."

## ORIGIN

Query Match	81.0%;	Score 162.8;	DB 12;	Length 302;
Best Local Similarity	96.0%;	Pred NO. 1	3e-39;	

Category	Count	Percentage
Marbles	107	3.1%
Conservative	0	0.0%
Misalliances	1	0.3%
Unlabeled	0	0.0%
Caps	0	0.0%

Qy	1	CCATGACGACGAGTATGTCCTTCAAAATCGAAGAGCTTTGACTGCTAAACGCTGAAGAGG	60
Db	129	CCATGACGACGAGGAGCCCTTCAAAACCGAAGAGCTTTAGACTGCTAAACGCTGAAGAGG	18
Qy	61	GGGAAACCTGTTATATTTATGGGGAAGATGCTGTTATATATGTAATCAATCCGAATCGT	12
Db	189	GGGAAACCTGTTATATTTATGGGGAAGATGCTGTTATATATGTTAATCAATCCGAATCGT	24
Qy	121	CACGTGAAAGTTAAAGAAATTCGAGATCGAATCAACAGTGAAGCAGAGAGCT	174
Db	249	CACGTGAAAGTTAAAGAAATTCGAGATCGAATCAACAGTGAAGCAGAGAGCT	302

QY	61	GGGAACCGTTATTTTATTTAGGGGAAGATCGTTATTATGTTAACTAATCCGAATCGT	120
Db	450	GGGAATCTGTTATTATTTTATGAGGAATCGTGTATTAGCTTATCAATCTGGAAATCGT	391
QY	121	CACGAGAAAGTAAAGAAATTCAGATCGAATCAACGTTAGACAGAGAGCTTGCAAA	180
Db	390	CACCGAANAAGTTTANNAGAAATTCAGATCGAATCAACATGACAGAGAGATTTCANAA	331
QY	181	CACGTGACCCCTGGGGCCTCCT	201
Db	330	CACCGAGACCTGGGAGCTCCT	310

RESULT 14	LOCUS	DEFINITION	ACCESION
A0725469/c	A0725469	608 bp DNA linear GSS 14-JUL-1999	HS_5403_A1_E02_T7A RPCT-11 Human Male BAC Library Homo sapiens genomic clone Plate=979 COL=3 ROW=I, DNA sequence.
			10725460

RESULT 15	
BG572445	
LOCUS	BG572445
DEFINITION	BG572445 790 bp mRNA linear EST 10-APR-2001
DESCRIPTION	602593490F01 NIH_MGC_79 Homo sapiens CDNA clone IMAGE:4720846 5', mRNA sequence.

ACCESSION	AQ/22405	GI:5485138
VERSION	AQ725469.1	
KEYWORDS	GSS.	
SOURCE	human.	

ACCESSION	BG572445	
VERSION	BG572445.1	GI:13580098
KEYWORDS	EST.	
SOURCE	human.	

ORGANISM	REFERENCE-AUTHORS
<i>Homo sapiens</i>	
Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:	
Mammalia: Eutheria: Primates: Catarrhini: Homnidae: Homo.	
1 (bases 1 to 608)	
Mahaitas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T.,	

ORGANISM  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 790)  
REFERENCE  
NIH-MGC <http://mgc.nci.nih.gov/>.

TITLE	Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 96 (17), 9739-9744 (1999)
MEDLINE	99380589
COMMENT	Contact: Mahairas GG, Wallace JC, Hood L

COMMENT

Contact: Robert Strausberg, Ph.D.  
Email: [cgaps-remail.nih.gov](mailto:cgaps-remail.nih.gov)  
Tissue Procurement: CLONTECH Laboratories, Inc.  
cDNA Library Preparation: CLONTECH Laboratories, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

High Throughput Sequencing Center  
University of Washington  
401 Queen Anne Avenue North, Seattle, WA 98109, USA  
Tel.: (206) 616-3618  
Fax: (206) 616-3887  
Email: jwallaceu.washington.edu  
Clones are derived from the human BAC library RPCI-11. For BAC library availability, please contact Pieter de Jong (pieterdejong.med.buffalo.edu). Clones may be purchased from BACPAC Resources ([http://bacpac.med.buffalo.edu/ordering\\_bac.htm](http://bacpac.med.buffalo.edu/ordering_bac.htm)) or from Research Genetics ([info@resgen.com](http://info@resgen.com)). BAC end Web Server: <http://www.htsc.washington.edu>  
Plate: 979 row: I column: 3  
Seq primer: T7  
Class: BAC ends  
High quality sequence stop: 608.

DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://Image.llnl.gov>  
 Plate: L1CM1576 row: c column: 23  
 High quality sequence stop: 624.  
 Location/Qualifiers  
     1..790  
         /organism="Homo sapiens"  
         /db\_xref="taxon:9606"  
         /clone="IMAGE:4720846"  
         /clone\_lib="NIH\_MGC\_79"  
         /lab\_host="DH10B (T1 phage-resistant)"  
         /note="Organ: Placenta; Vector: PDNR-LIB (Clontech);"  
         site\_1: sfil (ggcgctctggcc); site\_2: sfil (ggcgctatt  
         ); 5' and 3' adaptors were used in cloning as follows

```

FEATURES
SOURCE
1. .608
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="plate=979 Col=3 Row=1"
/clone_lib="RPCr-11 Human Male BAC Library"
/sex="Male"
/note="Vector: pBACE3.6; Site.1: EcoRI; Site.2: EcoRI; Male blood DNA was isolated with one randomly chosen donor and partially digested with a combination of EcoRI and EcoRII Methylase. Size selected DNA was cloned into the pBACE3.6 vector at EcoRI sites"
BASE COUNT
165 a 114 c 145 g 173 t 11 others
ORIGIN

```

BASE COUNT	210 a	202 c	164 g	214 t						
ORIGIN	Library. <sup>a</sup>									
Full-length clones and was constructed by Clontech Laboratories (Palo Alto, CA). Note: this is a NIH-MGC										
insert size 1.3 kb (range 0.5-4.0 kb). 15/15 colonies										
contained inserts by PCR. This library was enriched										
(where B = A, C, or G and N = A, C, G, or T). Average										
sequence: 5'-ATTCTAGAGCCGAGCGAGCGGCGCCACATG-dt(30)BN-3'										
adaptor sequence: 5'-CACGGCCATTGAGCC-3' and 3' adapt										
Query Match	75.88;	Score 152.4;	DB 12;	Length 790;						
Best Local Similarity	95.0%;	Pred. NO. 2.5e-36;								
Matches 190; Conservative	0;	Mismatches 6;	Indels 4;	Gaps						

Query Match	78.5%	Score 157.8	DB 17	Length 608
Best Local Similarity	89.1%	Pred. No. 5, 2e-38		
Matches 179	Conservative 0	Mismatches 21	Indels 1	Gaps 1

  

OY	1	CCTAGCAGCACTAGTCTCTTCAATGAGAGACTTGTAGACTTGCTAACCGCTGAAGAAGG	60
db	509	CTTAGCAGCACTAGTCTTTC-AAAATCGAAGAGCTTTTAACTGTCTTAACCTGAGAAAGAGA	451

QY	1	CCTGAGCAGCACTACTCCCTTCA - AATGAGAGAGCTTTAGACTGCTGTAACGGCTGAAGAAG	59
Db	381	CCTGACACCACTAGTCCCTTCACCAATGCAGAGCTTTAGACTCTGCTAACCGCTGAAGAAG	440
QY	60	GGGGAACCTGTTATTTTATGAGGGAGAGAATGCTGTTATTTATGTTAAATCAATCCGGAATCG	119
Db	441	GGGGAACCTGTTATTTTATTTTATGAGGGAGAGAATGCTGTTATTTATGTTAAATCAATCCGGAATCG	500



QY 120 TCACGTGAAAGTTA-AAGAAATTCGAGATCG--AATACACGTAGACAGAGAGCTTC 176  
 |||||  
 Db 501 TCACGTGAAAGTTACAGAAATTCGAGATCGCAATACACGCTAGAGCAGAGAGCTTC 560  
 |||||  
 QY 177 GAAACACTGGACCCCTGGGGC 196  
 |||||  
 Db 561 GAAACACTGGACCCCTGGGGC 580

Search completed: May 2, 2003, 16:21:36  
 Job time : 1116 secs



US-09-011-745-1

```

; Patent No. 6165715
; GENERAL INFORMATION:
; APPLICANT: Collins, Mary KL
; APPLICANT: Weiss, Robin A
; APPLICANT: Takeuchi, Yasuhiro
; APPLICANT: Cosset, Francois-Lolc
; TITLE OF INVENTION: Expression systems
; FILE REFERENCE: 09/011,745
; CURRENT APPLICATION NUMBER: US/09/011,745
; EARLIER FILING DATE: 1998-06-22
; EARLIER APPLICATION NUMBER: PCT/GB96/02061
; EARLIER FILING DATE: 1996-08-23
; EARLIER APPLICATION NUMBER: GB9517263.1
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 1
; LENGTH: 2518
; TYPE: DNA
; ORGANISM: RD114
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)
; OTHER INFORMATION: n is any nucleotide
US-09-011-745-1

Query Match          36.1%; Score 72.6; DB 4; Length 2518;
Best Local Similarity 60.3%; Pred. No. 9.3e-14;
Matches 120; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY 3 TACACAGTAGTCCTTCAAAATCGAAGAGCTTTAGACTGCTGACCGCTGAAGAAGGG 62
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1913 TACCTGAGTAGTCTCCAAAATAGGAGGAGCTGACCTACTACGCGCAGAACAGAG 1972

QY 63 GAACCTGTTATTTTAGGGGAAGAAATGCTATTATGTTATCAATCCGGAATGCTCA 122
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1973 GAATTTGTTTACCTTCAAGAAAATGCTGTTTATGCTACAGTCAGGAATGTGA 2032.

QY 123 CTGAGAAAGTTAAGAAATTCGAGATCGAATACAGCTAGAGCAGAGAGCTTCGAACA 182
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2033 GAAACAAAATTAAGAACCCCTACAGAAAGAAATTAACAAAACGACGGAAGCCTGGCAACA 2092

QY 183 CTGACCCCTGGGGCTCTCT 201
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2093 ACCCTCTCTGGACCGGCT 2111

RESULT 3
US-09-011-745-8
; Sequence 8, Application US/09011745
; Patent No. 6165715
; GENERAL INFORMATION:
; APPLICANT: Collins, Mary KL
; APPLICANT: Weiss, Robin A
; APPLICANT: Takeuchi, Yasuhiro
; APPLICANT: Cosset, Francois-Lolc
; TITLE OF INVENTION: Expression systems
; FILE REFERENCE: 09/011,745
; CURRENT APPLICATION NUMBER: US/09/011,745
; EARLIER FILING DATE: 1998-06-22
; EARLIER APPLICATION NUMBER: PCT/GB96/02061
; EARLIER FILING DATE: 1996-08-23
; EARLIER APPLICATION NUMBER: GB9517263.1
; NUMBER OF SEQ ID NOS: 29
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 8
; LENGTH: 5865
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Portion of
```

```

; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (3611)
; OTHER INFORMATION: n is any nucleotide
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (3612)
; OTHER INFORMATION: n is any nucleotide
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (3613)
; OTHER INFORMATION: n is any nucleotide
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (3614)
; OTHER INFORMATION: n is any nucleotide
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (3799)
; OTHER INFORMATION: n is any nucleotide
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (3800)
; OTHER INFORMATION: n is any nucleotide
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (3801)
; OTHER INFORMATION: n is any nucleotide
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (3802)
; OTHER INFORMATION: n is any nucleotide
US-09-011-745-8

Query Match          36.1%; Score 72.6; DB 4; Length 5865;
Best Local Similarity 60.3%; Pred. No. 1.2e-13;
Matches 120; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY 3 TACACAGTAGTCCTTCAAAATCGAAGAGCTTTAGACTGCTGACCGCTGAAGAAGGG 62
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2771 TACCTGAGTAGTCTCCAAAATAGGAGGAGCTGACCTACTACGCGCAGAACAGAG 2830

QY 63 GAACCTGTTATTTTAGGGGAAGAAATGCTATTATGTTATCAATCCGGAATGCTCA 122
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2831 GAATTTGTTTACCTTCAAGAAAATGCTGTTTATGCTACAGTCAGGAATGTGA 2890

QY 123 CTGAGAAAGTTAAGAAATTCGAGATCGAATACAGCTAGAGCAGAGAGCTTCGAACA 182
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2891 GAAACAAAATTAAGAACCCCTACAGAAAGAAATTAACAAAACGACGGAAGCCTGGCAACA 2950

QY 183 CTGACCCCTGGGGCTCTCT 201
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 2951 ACCCTCTCTGGACCGGCT 2969

RESULT 4
US-08-766-528-1
; Sequence 1, Application US/08766528
; Patent No. 6190861
; GENERAL INFORMATION:
; APPLICANT: Jay A. Fishman
; TITLE OF INVENTION: MOLECULAR SEQUENCE OF SWINE RETROVIRUS
; TITLE OF INVENTION: AND METHODS OF USE
; NUMBER OF SEQUENCES: 74
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LAHIVE & COCKFIELD, LLP
; STREET: 60 State Street
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02109-1875
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
```

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

Query Match	30.1%;	Score 60.6;	DB 4;	Length 8060;
Best Local Similarity	63.3%;	Pred. No. 8e-10;		
Matches 93;	Conservative 0;	Mismatches 54;	Indels 0;	Gaps 0;

Accession	Sequence	Position
OY 1	CTTACACACAGTAGTCCTTCAAAAATCGAAGAGCTTAGCTCTGCATACCGCGTGAAGAAG	60
Db	CTTATCTGAAGTAGTCCTACGAATAGAAGAGGGTTAGATTATTAATTTCTTAAAGAAGG	1659
OY 61	GGGAAACCGTGTATTTTATTAAGGGAGAAATCGCTTATATGTATATCAATCCGGAATGCT	120
Db	AGGATTATATGTAGCGCTTGAAAGAGGAAATGCTGTTTTTATGTGATCATTCAGGGGGCCAT	1719
OY 121	CACGTGAGAAAGTTAAAGAAATTCGAGA	147
Db	CAGAGCTCCATGAACCAAACTTAGGA	1746

RESULT 5  
US-08-007-282B-1  
; Sequence 1, Application US/08007282B  
; Patent No. 5403582

APPLICANT: NAZERIAN, KEYVAN  
APPLICANT: CALVERT, JAY G.  
APPLICANT: MITTER, RICHARD L.  
APPLICANT: YANAGIDA, NOBORU  
TITLE OF INVENTION: VACCINE COMPRISING FOWLPOX VIRUS  
TITLE OF INVENTION: RECOMBINANTS EXPRESSING THE ENVELOPE GLYCOPROTEIN OF  
TITLE OF INVENTION: AVIAN RETICULOENDOTHELIOSIS RETROVIRUS  
NUMBER OF SEQUENCES: 2  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BIRCH, STEWART, KOLASCH & BIRCH  
STREET: 8110 Gatehouse Road Suite 500 East  
CITY: Falls Church  
STATE: VA  
COUNTRY: USA  
ZIP: 22042  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/007,282B  
FILING DATE: 19930121  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: MURPHY JR., GERALD M.

```

: REGISTRATION NUMBER: 28.977
: REFERENCE/POCKET NUMBER: 1644-104P
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 703-205-8000
: TELEFAX: 703-205-8050
: TELEX: 248345
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1704 base pairs
: TYPE: NUCLEIC ACID
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: HYPOTHETICAL: NO
: ANTI-SENSE: NO
:
US-08-007-282B-1

```

Query Match	29.4%	Score 59;	DB 1;	Length 1704;
Best Local Similarity	62.6%;	Pred. NO. 1.6e-09;		
Matches 92;	Conservative 0;	Mismatches 55;	Indels 0;	Gaps 0;

Accession	Sequence	Position
Qy	1 CCTACACACAGTAGTCCTTCAAAATCGAAGAGCTTTACACTTGGCAACCGGTGAAAGAG	60
Db	1368 CCGTGTGAGGTGTCCTTACAAATAGAAAGAGGTACACTATTGATGCGCAACAAG	1427
Qy	61 GGGAAACCTGTTATTTTAGGGGAAGAAATGCTGTTATTATGTAACTCCGAAATGT	120
Db	1428 AGGAATATGTCGCGACATCCAGGAAAGAGTGTGTTTAACTGAACAAGTCGGGTATCT	1487
Qy	121 CACTGAGAAAGTTAAAGAAATTCGAGA	147
Db	1488 ACGTGACAAGATCCGAAAACCTCCAAGA	1514

RESULT 6  
US-08-766-528-3  
; Sequence 3, Application US/08766528  
; Patent No. 6190861

APPLICANT: Jay A. Fishman  
TITLE OF INVENTION: MOLECULAR SEQUENCE OF SWINE RETROVIRUS  
TITLE OF INVENTION: AND METHODS OF USE  
NUMBER OF SEQUENCES: 74  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: LAHIVE & COCKFIELD, LLP  
STREET: 60 State Street  
CITY: Boston  
STATE: Massachusetts  
COUNTRY: USA  
ZIP: 02109-1875  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentln Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/766,528  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/572,645  
FILING DATE: 14-DEC-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Louis Myers  
REGISTRATION NUMBER: 35,965  
REFERENCE/DOCKET NUMBER: MGP-038CP  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)227-7400  
TELEFAX: (617)227-5941  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 8132 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single

TOPOLOGY: linear  
MOLECULE TYPE: CDNA  
US-08-766-528-3

Query Match  
Best Local Similarity 29.4%; Score 59; DB 4; Length 8132;  
Matches 92; Conservative 0; Mismatches 55; Indels 0; Gaps 0;

QY 1 CCTAGCAGCAGTAGTCTCTTCAAAATCGAAGAGCTTTAGACTTGTCTAACCGCTGAAGAAGG 60  
DB 7134 CTTATCTGAAGTAGCTCTACAGAAATAGAGAAGGGTTAGATTATTTCTAAAGAAGG 7193  
QY 61 GCGAACCGTTATTTTATTTAGGGAAGATGCTTTATTTATCTATCATCCGGAATCGT 120  
DB 7194 AGGATTAGTGTACCTGGAAGAGGAACTGTTTATCTGATCATTCAGGGGCAT 7253  
QY 121 CACTGAGAAAGTTAAGAAATTCGAGA 147  
DB 7254 CAGAGACTCCATGAACAAGCTTAGAGA 7280

RESULT 7  
US-08-716-351A-5  
Sequence 5, Application US/08716351A  
Patent No. 6033905  
GENERAL INFORMATION:  
APPLICANT:  
TITLE OF INVENTION: Gibbon Ape Leukemia Virus-Based  
TITLE OF INVENTION: Retroviral Vectors  
NUMBER OF SEQUENCES: 5  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/716,351A  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US94/03784  
FILING DATE: 06-APR-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Bastian, Kevin L.  
REGISTRATION NUMBER: 34,774  
REFERENCE/DOCKET NUMBER: 15280-128-1PC  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 543-9600  
TELEFAX: (415) 543-5043  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 10970 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 1..10970  
OTHER INFORMATION: /standard\_name="p537 retroviral  
OTHER INFORMATION: vector"  
US-08-716-351A-5

Query Match  
Best Local Similarity 28.4%; Score 57; DB 3; Length 10970;  
Matches 108; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

QY 1 CTTAGCAGCAGTAGTCTCTTCAAAATCGAAGAGCTTTAGACTTGTCTAACCGCTGAAGAAGG 60  
DB 6894 CTTATCTGAAGTAGCTCTACAGAAATAGAGAAGGGTTAGATTATTTCTAAAGAAGG 6953  
QY 61 GCGAACCGTTATTTTATTTAGGGAAGATGCTTTATTTATCTATCATTCAGGGGCAT 120

DB 6954 AGCCCTGCGCGGCCCTTAAAGAGAGTGTCTTTTATGTAGACCACTCAGGTGCAGT 7013  
QY 121 CACTGAGAAAGTTAAGAAATTCGATCGAATCAACAGTAGACGAGAGCTTCGAAA 180  
DB 7014 ACAGACATCCATGAAAAAAGCTTTAAAGAAAGACTGATTAAGACAGTTAGAGCCCGAGAA 7073  
QY 181 CACTGGAGCCCTGG 193  
DB 7074 AAACCAAACTCG 7086

RESULT 8  
US-09-232-278A-5  
Sequence 5, Application US/09232278A  
Patent No. 6348196  
GENERAL INFORMATION:  
APPLICANT: AUDONNET et al.  
TITLE OF INVENTION: FELINE POLYNUCLEOTIDE VACCINE FORMULA  
FILE REFERENCE: 454313-2220  
CURRENT APPLICATION NUMBER: US/09/232,278A  
CURRENT FILING DATE: 1999-01-15  
NUMBER OF SEQ ID NOS: 30  
SOFTWARE: Patentin Ver. 2.1  
SEQ ID NO 5  
LENGTH: 1989  
TYPE: DNA  
ORGANISM: Feline leukemia virus  
US-09-232-278A-5

Query Match  
Best Local Similarity 28.2%; Score 56.6; DB 4; Length 1989;  
Matches 101; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

QY 1 CCTAGCAGCAGTAGTCTCTTCAAAATCGAAGAGCTTTAGACTTGTCTAACCGCTGAAGAAGG 60  
DB 1584 CCTCTCGAGGTACTCTTCAAAATAGACGCGGCTGATATCTGTTCTTCAAAAAGG 1643  
QY 61 GCGAACCGTTATTTTATTTAGGGAAGATGCTTTATTTATGTATCATCCGGAATCGT 120  
DB 1644 AGGCTCTGTGCGCCCTTAAAGAGAGATGCTGTTCTATGACATACACCGGACTCGT 1703  
QY 121 CACTGAGAAAGTTAAGAAATTCGAGATCGAATTCACAGTAGACGAGAGAGCTT 175  
DB 1704 CAGAGACAATATGCTAAATTAAGAGAAAGACTGAACAGCAGACACACTGTTT 1758

RESULT 9  
US-08-105-483-310  
Sequence 310, Application US/08105483  
Patent No. 5494807  
GENERAL INFORMATION:  
APPLICANT: Paolletti, Enzo  
TITLE OF INVENTION: GENETICALLY ENGINEERED VACCINE  
TITLE OF INVENTION: STRAIN  
NUMBER OF SEQUENCES: 462  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Curtis, Morris & Safford  
ADDRESSEE: c/o William S. Frommer  
STREET: 530 Fifth Avenue  
CITY: New York  
STATE: NY  
COUNTRY: USA  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/105,483  
FILING DATE: 12-AUG-1993  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/847,951  
FILING DATE: 06-MAR-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Frommer, William S.  
REGISTRATION NUMBER: 25,506  
REFERENCE/DOCKET NUMBER: 454310-2400  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 840-3333  
TELEFAX: (212) 840-0712  
INFORMATION FOR SEQ ID NO: 310:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2499 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-105-483-310

Query Match 28.2%; Score 56.6; DB 1; Length 2499;  
Best Local Similarity 57.7%; Pred. No. 1e-08;  
Matches 101; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

QY 1 CCTAGACAGTAGTCCTCAAAATCGAAGAGCTTGAAGCTTGAACCGCTGAAGAG 60  
DB 1608 CCTCTCCGAGTAGCTTCAAAATAGACGGCGCTAGATATCTCTTACAAAGG 1667

QY 61 GGAACCTGTTATTTTATGAGGAGATGCTGTTATTTATGTAATCAATCCGAAATCGT 120  
DB 1668 AGGGCTCTGTGCCCTTAAAGGAGAAATGCTGCTTATGAGATCAGACCGGACTCGT 1727

QY 121 CACTGAGAAAGTTAAAGAAATTCGAGATCGAATACAGCTAGAGCAGAGAGCTT 175  
DB 1728 CAGAGACAATATGCTAAATTAAGAGAAAGACTGAACAGCGACACAACTGTTT 1782

RESULT 10  
US-08-709-209-310  
Sequence 310, Application US/08709209  
Patent No. 5762938  
GENERAL INFORMATION:  
APPLICANT: Paoletti, Enzo  
TITLE OF INVENTION: GENETICALLY ENGINEERED VACCINE  
TITLE OF INVENTION: STRAIN  
NUMBER OF SEQUENCES: 462  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Curtis, Morris & Safford  
STREET: 530 Fifth Avenue  
CITY: New York  
STATE: NY  
COUNTRY: USA  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/709,209  
FILING DATE: 21-AUG-1996  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/105,483  
FILING DATE: 12-AUG-1993  
APPLICATION NUMBER: US 07/847,951  
FILING DATE: 06-MAR-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Frommer, William S.  
REGISTRATION NUMBER: 25,506  
REFERENCE/DOCKET NUMBER: 454310-2400  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 840-3333  
TELEFAX: (212) 840-0712  
INFORMATION FOR SEQ ID NO: 310:

SEQUENCE CHARACTERISTICS:  
LENGTH: 2499 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-709-209-310

Query Match 28.2%; Score 56.6; DB 1; Length 2499;  
Best Local Similarity 57.7%; Pred. No. 1e-08;  
Matches 101; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

QY 1 CCTAGACAGTAGTCCTCAAAATCGAAGAGCTTGAAGCTTGAACCGCTGAAGAG 60  
DB 1608 CCTCTCCGAGTAGCTTCAAAATAGACGGCGCTAGATATCTCTTACAAAGG 1667

QY 61 GGAACCTGTTATTTTATGAGGAGATGCTGTTATTTATGTAATCAATCCGAAATCGT 120  
DB 1668 AGGGCTCTGTGCCCTTAAAGGAGAAATGCTGCTTATGAGATCAGACCGGACTCGT 1727

QY 121 CACTGAGAAAGTTAAAGAAATTCGAGATCGAATACAGCTAGAGCAGAGAGCTT 175  
DB 1728 CAGAGACAATATGCTAAATTAAGAGAAAGACTGAACAGCGACACAACTGTTT 1782

RESULT 11  
US-08-458-101-310  
Sequence 310, Application US/08458101  
Patent No. 5766599  
GENERAL INFORMATION:  
APPLICANT: Paoletti, Enzo  
TITLE OF INVENTION: GENETICALLY ENGINEERED VACCINE  
TITLE OF INVENTION: STRAIN  
NUMBER OF SEQUENCES: 467  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Curtis, Morris & Safford  
STREET: 530 Fifth Avenue  
CITY: New York  
STATE: NY  
COUNTRY: USA  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/458,101  
FILING DATE: 01-JUN-1995  
CLASSIFICATION: 424  
ATTORNEY/AGENT INFORMATION:  
NAME: Frommer, William S.  
REGISTRATION NUMBER: 25,506  
REFERENCE/DOCKET NUMBER: 454310-2740  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 840-3333  
TELEFAX: (212) 840-0712  
INFORMATION FOR SEQ ID NO: 310:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2499 base pairs  
TYPE: nucleic acid

STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-458-101-310

Query Match 28.2%; Score 56.6; DB 1; Length 2499;  
Best Local Similarity 57.7%; Pred. No. 1e-08;  
Matches 101; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

QY 1 CCTACGACGATGCTCTCCAAATCGAAGACCTTTAGACTGCTTACCGCTGAAGAGG 60  
DB 1608 CCTCTCCGAGTAGTCTTACCAAAATAGACGGGCGCTAGATATCTGTTCTTACAAAAGG 1667  
QY 61 GGAACCTGTTATTTTATGAGGAATGCTTATATGTTATTCATCCGGAATGCT 120  
DB 1668 AGGCTCTGTGCGCCCTTAAGGAAGATGCTCTTATGCAATCACCGGACTGCT 1727  
QY 121 CACTAGAAAGTTAAAGAAATTCGAGATCGAATACAGCTAGAGGAGGACTT 175  
DB 1728 CAGAGACAATATGCTTAATTAAGAGAAAGACTGAACGAGCAACAACGTGTT 1782

## RESULT 12

US-08-691-563C-58  
Sequence 58, Application US/08691563C  
Patent No. 6001987

## GENERAL INFORMATION:

APPLICANT: Hervé PERRON  
APPLICANT: Frédéric BESEME  
APPLICANT: Frédéric BÉDIN  
APPLICANT: Glaucia PARANHOS-BACCALA  
APPLICANT: Florence KOMURIAN-PRADEL  
APPLICANT: Colette JOLIVET  
APPLICANT: Bernard MANDRAND  
TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS  
TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTIC  
NUMBER OF INVENTION: THERAPEUTIC PURPOSES  
NUMBER OF SEQUENCES: 92  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Oliff & Berridge  
STREET: 700 South Washington Street, Suite 300  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22314

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/691,563C  
FILING DATE: 02-AUG-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Berridge, William P.  
REGISTRATION NUMBER: 30,024  
REFERENCE/DOCKET NUMBER: WPB 38588  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703-836-6400  
TELEFAX: 703-836-2787

## INFORMATION FOR SEQ ID NO: 58:

SEQUENCE CHARACTERISTICS:  
LENGTH: 1722 base pairs  
TYPE: nucleotide  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
US-08-691-563C-58

Query Match 26.9%; Score 54; DB 3; Length 1722;  
Best Local Similarity 66.1%; Pred. No. 6e-08;  
Matches 78; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 3 TAGACGACGATGCTCTCCAAATCGAAGACCTTTAGACTGCTTAACCGCTGAAGAGGG 62

DB 530 TGGCAGCATGACTCTCCAAACCCGAGGCCACACCTCTCTCACTGTGAGAAAGAG 589

QY 63 GAACCTGTTATTTTATGAGGAATGCTGTTATATGTTAATCAATCCGAATGCT 120  
DB 590 GACTCTGCACCTTCTTAAGGGAAGAGTGTGTTTACACTACCACTGAGGATAGT 647

## RESULT 13

US-08-691-563C-46  
Sequence 46, Application US/08691563C  
Patent No. 6001987

## GENERAL INFORMATION:

APPLICANT: Hervé PERRON  
APPLICANT: Frédéric BESEME  
APPLICANT: Frédéric BÉDIN  
APPLICANT: Glaucia PARANHOS-BACCALA  
APPLICANT: Florence KOMURIAN-PRADEL  
APPLICANT: Colette JOLIVET  
APPLICANT: Bernard MANDRAND  
TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS  
TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYLACTIC  
NUMBER OF INVENTION: THERAPEUTIC PURPOSES  
NUMBER OF SEQUENCES: 92  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Oliff & Berridge  
STREET: 700 South Washington Street, Suite 300  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: U.S.A.  
ZIP: 22314

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/691,563C  
FILING DATE: 02-AUG-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Berridge, William P.  
REGISTRATION NUMBER: 30,024  
REFERENCE/DOCKET NUMBER: WPB 38588  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703-836-6400  
TELEFAX: 703-836-2787  
INFORMATION FOR SEQ ID NO: 46:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1859 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
US-08-691-563C-46

Query Match 26.9%; Score 54; DB 3; Length 1859;  
Best Local Similarity 66.1%; Pred. No. 6.1e-08;  
Matches 78; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

QY 3 TAGACGACGATGCTCTCCAAATCGAAGACCTTTAGACTGCTTAACCGCTGAAGAGGG 62  
DB 1180 TGGCAGCATGACTCTCCAAACCGGAGGCCCTAGACTCTCTGAGAGAAAGAG 1239

QY 63 GAACCTGTTATTTTATGAGGAATGCTGTTATATGTTAATCAATCCGAATGCT 120  
DB 1240 GACTCTGCACCTTCTTAAGGGAAGAGTGTGTTTACACTACCACTGAGGATAGT 1297

## RESULT 14

US-09-075-272-1  
Sequence 1, Application US/09075272  
Patent No. 6136598

## GENERAL INFORMATION:



APPLICANT: MILLER, A. DUSTY  
APPLICANT: WOLGAST, GREG  
APPLICANT: BONHAM, LYNN  
TITLE OF INVENTION: MUS DUNNI ENDOGENOUS RETROVIRAL  
TITLE OF INVENTION: PACKAGING CELL LINES  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: TOWNSEND and TOWNSEND and CREW LLP  
STREET: Two Embarcadero Center, 8th floor  
City: San Francisco  
STATE: California  
COUNTRY: U.S.A.  
ZIP: 94111  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/075,272  
FILING DATE: 08-MAY-1998  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 60/046,140  
FILING DATE: 09-MAY-1997  
ATTORNEY/AGENT INFORMATION:  
NAME: POOR, Brian W.  
REGISTRATION NUMBER: 32,928  
REFERENCE/DOCKET NUMBER: 14538A-003710  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 467-9600  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ. ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 8655 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
US-09-075-272-1

Query Match 25.4%; Score 51; DB 3; Length 8655;  
Best Local Similarity 58.1%; Pred. No. 8.6e-07;  
Matches 90; Conservative 0; Mismatches 65; Indels 0; Gaps 0;  
QY 2 CTACGACAGTAGTCTTCAAAATCGAAGAGCTTTAGACTTGTACCGCTGAAGAGGG 61  
DB 7374 CTGCTGAGGTAGTACTGCAAAATAGGAGAGATTAGATTGTTCTTAAGAAGA 7433  
QY 62 GGAACCTGTTATTTTAAAGGGAAGATGCTTATATGTTATCAATCCGGAATGCTC 121  
DB 7434 GGAATTGCTGCTGCTCAAGAGAGAAATGTTGTTCTATGTCGACACTGGGAGTGATC 7493  
QY 122 ACTGAGAAAGTTAAAGAAATTCGATCGAATACA 156  
DB 7494 AAGATTCTATGGCCAAACTTAGAGAACGCTTACA 7528

RESULT 15  
US-08-552-369-19  
Sequence 19, Application US/08552369  
Patent No. 6241989  
GENERAL INFORMATION:  
APPLICANT: Scott, Fred W.  
APPLICANT: Ngichabe, Christopher K.  
APPLICANT: Hu, Liangbiao  
TITLE OF INVENTION: Recombinant Multivalent Viral Vaccine  
NUMBER OF SEQUENCES: 19  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Hodgson, Russ, Andrews, Woods & Goodyear  
STREET: 1800 One Met Plaza  
City: Buffalo  
STATE: New York

COUNTRY: United States  
ZIP: 14203-2391  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Kb storage  
COMPUTER: IBM compatible  
OPERATING SYSTEM: MS-DOS/ Microsoft Windows 3.1  
SOFTWARE: Wordperfect  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/552,369  
FILING DATE:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/190,789  
FILING DATE: 01/27/1994  
APPLICATION NUMBER: 07/726,609  
FILING DATE: 07/09/1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Nelson, M. Bud  
REGISTRATION NUMBER: 35,300  
REFERENCE/DOCKET NUMBER: 18617.0016  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (716) 849-0349  
TELEFAX: (716) 849-0349  
INFORMATION FOR SEQ. ID NO: 19:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1979 nucleotides  
TYPE: nucleic acid  
STRANDEDNESS: double-stranded  
TOPOLOGY: linear  
MOLECULE TYPE: DNA  
HYPOTHETICAL: Yes  
ORIGINAL SOURCE:  
ORGANISM: feline leukemia virus  
FEATURE:  
LOCATION: env gene region  
US-08-552-369-19

Query Match 25.3%; Score 50.8; DB 4; Length 1979;  
Best Local Similarity 57.6%; Pred. No. 6.4e-07;  
Matches 91; Conservative 0; Mismatches 67; Indels 0; Gaps 0;  
QY 1 CCTAGCAGCAGTAGTCTTCAAAATCGAAGAGCTTTAGACTTGTACCGCTGAAGAGG 60  
DB 1574 CCTTTCGAGAGTAGTCTTACAAACAGCGGGCTGATATTTCTTACAAAGGG 1633  
QY 61 GGAACCTGTTATTTTAAAGGGAAGATGCTTATATGTTATCAATCCGGAATGCT 120  
DB 1634 AGGGCTCTGTGCCCATTTGAAGAAATGTTGCTTCTATGCGGATCACACCGGACTGCT 1693  
QY 121 CACTGAGAAAGTTAAAGAAATTCGATCGAATCAAC 158  
DB 1694 CCGAGACAAATATGGCCAAATTAAAGAGAAAGCTAAAC 1731

Search completed: May 2, 2003, 16:24:15  
Job time : 53.3333 secs



GenCore version 5.1.5  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 2, 2003, 16:21:41 ; Search time 63.333 Seconds

(Without alignments)  
3750.092 Million cell updates/sec

Title: US-09-719-554-3\_COPY\_9000\_9200

Perfect score: 201  
Sequence: 1 cctagcagcagtagtctctc.....actgacctggggcctctc 201

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 746064 seqs, 590810554 residues

Total number of hits satisfying chosen parameters: 1492128

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published\_Applications\_NA:\*  
1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/1/pubpna/PCF\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/ptodata/1/pubpna/PCFUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*  
10: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:\*  
12: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq:\*  
13: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:\*  
14: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	201	100.0	2930	10 US-09-902-535-1	Sequence 1, Appl1
2	201	100.0	2946	9 US-10-114-893-134	Sequence 134, App
3	183.4	91.2	1481	7 US-08-979-847-105	Sequence 105, App
4	178.6	88.9	1894	10 US-09-864-761-4444	Sequence 4444, App
5	163.8	81.5	1329	7 US-08-979-847-108	Sequence 108, App
6	127.6	48.5	426	10 US-09-864-761-14030	Sequence 14030, A
7	97.4	48.5	152	10 US-09-864-761-30590	Sequence 30590, A
8	89.4	44.5	520	10 US-09-864-761-7310	Sequence 7310, App
9	84.4	42.0	559	10 US-09-864-761-7501	Sequence 7501, App
10	73.4	36.5	960	9 US-09-764-891-283	Sequence 283, App
11	73.4	36.5	5434	9 US-10-001-835-58	Sequence 58, Appl1
12	71.8	35.7	310	10 US-09-864-761-24000	Sequence 24000, A
13	71.8	35.7	556	10 US-09-864-761-7269	Sequence 7269, App
14	65.8	32.7	1783	10 US-09-729-674-125	Sequence 125, App
15	65.8	32.7	2342	10 US-09-739-254-12	Sequence 12, Appl1
16	65.8	32.7	2342	10 US-09-904-615-12	Sequence 12, Appl1
17	65.8	32.7	5443	9 US-09-764-891-9093	Sequence 9093, App
18	64.6	32.1	143068	10 US-09-967-768A-316	Sequence 316, App
19	58	28.9	8928	9 US-09-764-891-6916	Sequence 6916, App

20	58	28.9	8928	9 US-09-764-891-9705	Sequence 9705, App
21	55	27.4	551	10 US-09-864-761-6686	Sequence 6686, App
22	54	26.9	1722	7 US-08-979-847-54	Sequence 54, Appl1
23	54	26.9	1859	7 US-08-979-847-42	Sequence 42, Appl1
24	49.4	24.6	1493	10 US-09-851-859A-24	Sequence 24, Appl1
25	49.4	24.6	2000	10 US-09-851-859A-25	Sequence 25, Appl1
26	49.4	24.6	2000	10 US-09-851-859A-30	Sequence 30, Appl1
27	48	23.9	792	10 US-09-864-761-21192	Sequence 21192, A
28	47.8	23.8	5130	10 US-09-887-006-9	Sequence 9, Appl1
29	47.8	23.8	8332	10 US-09-006-298-1	Sequence 1, Appl1
30	47.4	23.6	481	9 US-09-918-995-11010	Sequence 11010, A
31	47	23.4	3027	10 US-09-974-298-177	Sequence 177, App
32	45.8	22.8	532	10 US-09-998-598-421	Sequence 421, App
33	45	22.4	505	10 US-09-998-598-384	Sequence 384, App
34	45	22.4	529	10 US-09-920-300A-997	Sequence 997, App
35	45	22.4	529	10 US-10-033-928-997	Sequence 997, App
36	45	22.4	599	10 US-09-998-598-533	Sequence 533, App
37	44.4	22.1	395	10 US-09-864-761-31860	Sequence 31860, A
38	44.4	22.1	579	10 US-09-864-761-15338	Sequence 15338, A
39	44.2	22.0	454	10 US-09-998-598-471	Sequence 471, App
40	43.8	21.8	491	10 US-09-864-761-5518	Sequence 5518, App
41	43.6	21.7	539	10 US-09-919-580-156	Sequence 156, App
42	43.4	21.6	299	10 US-09-998-598-571	Sequence 571, App
43	43.4	21.6	310	10 US-09-998-598-490	Sequence 490, App
44	43.4	21.6	405	10 US-09-998-598-891	Sequence 891, App
45	43.4	21.6	459	10 US-09-998-598-552	Sequence 552, App

## ALIGNMENTS

RESULT 1  
US-09-902-535-1  
; Sequence 1, Application US/09902535  
; Patent No. US20020102530A1  
; GENERAL INFORMATION:  
; APPLICANT: Keith, Jr., James C.  
; APPLICANT: McCoy, John M.  
; TITLE OF INVENTION: Methods and compositions for diagnosing  
; TITLE OF INVENTION: and treating preeclampsia and gestational trophoblast  
; FILE REFERENCE: GIN-600684  
; CURRENT APPLICATION NUMBER: US/09/902,535  
; CURRENT FILING DATE: 2001-07-09  
; PRIOR APPLICATION NUMBER: 60/216,657  
; PRIOR FILING DATE: 2000-07-06  
; NUMBER OF SEQ ID NOS: 5  
; SOFTWARE: FASTSEQ for Windows Version 4.0  
; SEQ ID NO 1  
; LENGTH: 2930  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (930)...(2546)  
US-09-902-535-1  
Query Match 100.0%; Score 201; DB 10; Length 2930;  
Best Local Similarity 100.0%; Pred. No. 2.9e-52;  
Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
OY 1 CCTACAGCAGTAGTCCTTCAAAATCGAAGCCTTAGACTTGTCAACCGCTGAAGAG 60  
DB 2051 CCTACAGCAGTAGTCCTTCAAAATCGAAGCCTTAGACTTGTCAACCGCTGAAGAG 2110  
OY 61 GGGACCGTGTATTTTAAAGGAGAGTGTATATGTATATCAATCGGAATCGT 120  
DB 2111 GGGACCGTGTATTTTAAAGGAGAGTGTATATGTATATCAATCGGAATCGT 2170  
OY 121 CACTGAGAAAGTTAAAGAAATTCGAGATCGAATACAGCTAGAGAGAGAGCTTCGAAA 180  
DB 2171 CACTGAGAAAGTTAAAGAAATTCGAGATCGAATACAGCTAGAGAGAGAGCTTCGAAA 2230

OY 181 CACTGACCGTGGGCTCTCT 201  
|||||  
Db 2231 CACTGACCGTGGGCTCTCT 2251

## RESULT 2

US-10-114-893-134  
; Sequence 134, Application US/10114893  
; Publication No. US20020193567A1  
; GENERAL INFORMATION:  
; APPLICANT: Jacobs, Kenneth  
; APPLICANT: McCoy, John M.  
; APPLICANT: Lavallee, Edward R.  
; APPLICANT: Collins-Racie, Lisa A.  
; APPLICANT: Evans, Cheryl  
; APPLICANT: Merberg, David  
; APPLICANT: Treacy, Maurice  
; APPLICANT: Bowman, Michael R.  
; APPLICANT: Spaulding, Vikki  
; APPLICANT: Carlin-Duckett, McKeough  
; APPLICANT: Kelleher, Kerry S.  
; APPLICANT: Genetics Institute, Inc.  
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM  
; FILE REFERENCE: GI 6000-10A  
; CURRENT APPLICATION NUMBER: US/10/114,893  
; EARLIER FILING DATE: 2002-04-02  
; EARLIER APPLICATION NUMBER: 09/413,232  
; NUMBER OF SEQ ID NOS: 321  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 134  
; LENGTH: 2946  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-10-114-893-134

Query Match 100.0%; Score 201; DB 9; Length 2946;  
Best Local Similarity 100.0%; Pred. No. 2.9e-52;  
Matches 201; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 CCTAGCAGCAGTACTCCTCAAAATCGAAGAGCTTTAGACTTGTAAACCGCTGAAGAGG 60  
|||||  
Db 2049 CCTAGCAGCAGTACTCCTCAAAATCGAAGAGCTTTAGACTTGTAAACCGCTGAAGAGG 2108  
|||||  
OY 61 GGAACCTGTTATTTTATTTAGGGAAGATGCTTTATTTATGTTAATCAATCCGAATCGT 120  
|||||  
Db 2109 GGAACCTGTTATTTTATTTAGGGAAGATGCTTTATTTATGTTAATCAATCCGAATCGT 2168  
|||||  
OY 121 CACTGAGAAAGTTAAAGAAATTCGAGATCGAATCAACGTRAGACAGAGAGCTTCGAAA 180  
|||||  
Db 2169 CACTGAGAAAGTTAAAGAAATTCGAGATCGAATCAACGTRAGACAGAGAGCTTCGAAA 2228  
|||||  
OY 181 CACTGACCGTGGGCTCTCT 201  
|||||  
Db 2229 CACTGACCGTGGGCTCTCT 2249

## RESULT 3

US-08-979-847-105  
; Sequence 105, Application US/08979847  
; Publication No. US20030039664A1  
; GENERAL INFORMATION:  
; APPLICANT: PERRON, HERVE  
; APPLICANT: BESEME, FREDERIC  
; APPLICANT: BEDIN, FREDERIC  
; APPLICANT: PARANHOS-BACCALA, GLAUCIA  
; APPLICANT: KOMURIAN-PRADEL, FLORENCE  
; APPLICANT: JOLIVET-REYNAUD, COLETTE  
; APPLICANT: MANDRAND, BERNARD  
; APPLICANT: GARSON, JEREMY  
; APPLICANT: TURE, PHILIP  
; TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS

;; TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPH  
;; TITLE OF INVENTION: THERAPEUTIC PURPOSES  
;; NUMBER OF SEQUENCES: 210  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: OLIVE & BERRIDGE, PLC  
;; STREET: P. O. BOX 19928  
;; CITY: ALEXANDRIA  
;; STATE: VA  
;; COUNTRY: USA  
;; ZIP: 22320

## COMPUTER READABLE FORM:

;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.30  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/979,847  
;; FILING DATE: 26-NOV-1997  
;; CLASSIFICATION: 435  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: BERRIDGE, WILLIAM P.  
;; REGISTRATION NUMBER: 30,024  
;; REFERENCE/DOCKET NUMBER: WPB 39046A  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: 703-836-6400  
;; TELEFAX: 703-836-2787  
;; INFORMATION FOR SEQ ID NO: 105:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 1481 base pairs  
;; TYPE: nucleic acid  
;; STRANDEDNESS: single  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: cDNA  
US-08-979-847-105

Query Match 91.2%; Score 183.4; DB 7; Length 1481;  
Best Local Similarity 94.5%; Pred. No. 6.1e-47;  
Matches 190; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

OY 1 CCTAGCAGCAGTACTCCTCAAAATCGAAGAGCTTTAGACTTGTAAACCGCTGAAGAGG 60  
|||||  
Db 1122 CCTAGCAGCAGTACTCCTCAAAATCGAAGAGCTTTAGACTTGTAAACCGCTGAAGAGG 1181  
|||||  
OY 61 GGAACCTGTTATTTTATTTAGGGAAGATGCTTTATTTATGTTAATCAATCCGAATCGT 120  
|||||  
Db 1182 GGAACCTGTTATTTTATTTAGGGAAGATGCTTTATTTATGTTAATCAATCCGAATCGT 1241  
|||||  
OY 121 CACTGAGAAAGTTAAAGAAATTCGAGATCGAATCAACGTRAGACAGAGAGCTTCGAAA 180  
|||||  
Db 1242 CACTGAGAAAGTTAAAGAAATTCGAGATCGAATCAACGTRAGACAGAGAGCTTCGAAA 1301  
|||||  
OY 181 CACTGACCGTGGGCTCTCT 201  
|||||  
Db 1302 CACTGACCGTGGGCTCTCT 1322

## RESULT 4

US-09-864-761-4444  
; Sequence 4444, Application US/09864761  
; Patent No. US20020048763A1  
; GENERAL INFORMATION:  
; APPLICANT: Penn, Sharon G.  
; APPLICANT: Rank, David R.  
; APPLICANT: Hanzel, David K.  
; APPLICANT: Chen, Wensheng  
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FO  
; FILE REFERENCE: Aeomica-X-1  
; CURRENT APPLICATION NUMBER: US/09/864,761  
; CURRENT FILING DATE: 2001-05-23  
; PRIOR APPLICATION NUMBER: US 60/180,312  
; PRIOR FILING DATE: 2000-02-04  
; PRIOR APPLICATION NUMBER: US 60/207,456

PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: US 09/632,366  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: GB 24263.6  
PRIOR FILING DATE: 2000-10-04  
PRIOR APPLICATION NUMBER: US 60/236,359  
PRIOR FILING DATE: 2000-09-27  
PRIOR APPLICATION NUMBER: PCT/US01/00666  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00667  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00664  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00669  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00665  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00668  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00663  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00662  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00661  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: PCT/US01/00670  
PRIOR FILING DATE: 2001-01-30  
PRIOR APPLICATION NUMBER: US 60/234,687  
PRIOR FILING DATE: 2000-09-21  
PRIOR APPLICATION NUMBER: US 09/608,408  
PRIOR FILING DATE: 2000-06-30  
PRIOR APPLICATION NUMBER: US 09/774,203  
PRIOR FILING DATE: 2001-01-29  
NUMBER OF SEQ ID NOS: 49117  
SOFTWARE: Annomax Sequence Listing Engine vers. 1.1  
SEQ ID NO 4444  
LENGTH: 1894  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AC002346.1  
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 20  
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 6  
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 5.4  
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 5.9  
OTHER INFORMATION: EXPRESSED IN PETAL LIVER, SIGNAL = 4  
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 6.7  
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 6.2  
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 6.3  
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 3.9  
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 7.2  
US-09-864-761-4444

Query Match 88.9%; Score 178.6; DB 10; Length 1894;  
Best Local Similarity 93.0%; Pred. No. 2.1e-45;  
Matches 187; Conservative 0; Mismatches 14; Indels 0; Gaps 0;

DB 1 CCGAGCAGCGTAGCTCTTAAATGCAAGAGCTTTAGACTTGTCAACCGCTGAAGAGG 60  
DB 943 CCGAGCAGCGTAGCTCTTAAATGCAAGAGCTTTAGACTTGTCAACCGCTGAAGAGG 1002  
QY 61 GGGACCTGTTATTTTAAAGGAGAGAGCTTTATTTATTTATTCATCCGGAATCGT 120  
DB 1003 GGGACCTGTTATTTTAAAGGAGAGAGCTTTATTTATTTATTCATCCGGAATCAT 1062  
QY 121 CACTGAGAAAGTTAAAGAAATTCGAGATCGAATACACGTAAGACAGAGAGCTTCGAAA 180  
DB 1063 CACCGAGAAAGTTAAAGAAATTCGAGATCGAATACACGTAAGACAGAGAGCTTCGAAA 1122  
QY 181 CACTGAGACCTGGGCGCTCT 201  
DB 1123 CACTGAGACCTGGGCGCTCT 1143

RESULT 5  
US-08-979-847-108  
Sequence 108, Application US/08979847  
Publication No. US20030039664A1

GENERAL INFORMATION:

APPLICANT: PERRON, HERVE  
APPLICANT: BESEME, FREDERIC  
APPLICANT: BEDIN, FREDERIC  
APPLICANT: PARANHOS-BACCALA, GLAUCIA  
APPLICANT: KOMURIAN-PRADEL, FLORENCE  
APPLICANT: JOLIVET-REYNAUD, COLETTE  
APPLICANT: MANDRAND, BERNARD  
APPLICANT: GARSON, JEREMY

TITLE OF INVENTION: VIRAL MATERIAL AND NUCLEOTIDE FRAGMENTS  
TITLE OF INVENTION: ASSOCIATED WITH MULTIPLE SCLEROSIS, FOR DIAGNOSTIC, PROPHYL  
NUMBER OF SEQUENCES: 210  
CORRESPONDENCE ADDRESS:

ADDRESSEE: OLIVE & BERRIDGE, PLC  
STREET: P.O. BOX 19928  
CITY: ALEXANDRIA  
STATE: VA  
COUNTRY: USA

ZIP: 22320

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/979,847

FILING DATE: 26-NOV-1997

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: BERRIDGE, WILLIAM P.

REGISTRATION NUMBER: 30,024

REFERENCE/DOCKET NUMBER: WPB 39046A

TELECOMMUNICATION INFORMATION:

TELEPHONE: 703-836-6400

TELEFAX: 703-836-2787

INFORMATION FOR SEQ ID NOS: 108:

SEQUENCE CHARACTERISTICS:

LENGTH: 1329 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: cDNA

US-08-979-847-108

Query Match 81.5%; Score 163.8; DB 7; Length 1329;  
Best Local Similarity 93.4%; Pred. No. 6.9e-41;  
Matches 171; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 19 TCAAAATCGAAGAGCTTACACTGTAACCGCGTAAGAGAGGGGAACCGTTATTTT 78  
DB 1 TCAAAATCGAAGAGCTTACACTGTAACCGCGTAAGAGAGGGGAACCGTTATTTT 60  
QY 79 AGGGGAAGATGCTGTTATTTATTTATTTATTCATCCGGAATCGTCACTGAGAAAGTTAAGA 138  
DB 61 AGGGGAAGATGCTGTTATTTATTTATTTATTCATCCGGAATCATTAAGTAAGTAAGA 120  
QY 139 AATTGAGATCGAATACACGTAAGAGAGAGCTTCGAAACACTGAGACCTGGGGCTT 198  
DB 121 AATTGAGATCGAATACACGTAAGAGAGAGCTTCGAAACACTGAGACCTGGGGCTT 180  
QY 199 CCT 201  
DB 181 CCT 183

RESULT 6

```
US-09-864-761-14030
Sequence 14030, Application US/09864761
Patent No. US20020048763A1

GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng

TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aecomica-X-1
CURRENT FILING DATE: 2001-05-23
PRIORITY FILING DATE: 2001-05-23
PRIORITY FILING DATE: 2000-02-04
PRIORITY FILING DATE: 2000-05-26
PRIORITY FILING DATE: 2000-08-03
PRIORITY FILING DATE: GB 24263.6
PRIORITY FILING DATE: 2000-10-04
PRIORITY APPLICATION NUMBER: US 60/236,359
PRIORITY FILING DATE: 2000-09-27
PRIORITY APPLICATION NUMBER: PCT/US01/00666
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00667
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00664
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00669
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00665
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00668
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00663
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00662
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00661
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00670
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: US 60/234,687
PRIORITY FILING DATE: 2000-09-21
PRIORITY APPLICATION NUMBER: US 09/608,408
PRIORITY FILING DATE: 2000-06-30
PRIORITY APPLICATION NUMBER: US 09/774,203
PRIORITY FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117

SOFTWARE: Annomax Sequence Listing Engine vers. 1.1
SEQ ID NO 14030
LENGTH: 426
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AI022067.1
OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 0.95
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.1
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.3
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.2
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 7.1
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.6
OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.2
OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.1
US-09-864-761-14030
```

Query Match	63.5%	Score 127.6	DB 10	Length 426
Best Local Similarity	80.6%	Pred. No. 6.9e-30		
Matches 187	Conservative 0	Mismatches 14	Indels 31	Gaps 2

0y 1 CCTACGACAGTACGTCCTTCAAAATCGAAGAGCTTACACTTGCTTACCGCTGAAGAAGG 60  
 |||||||

Db 36 CCGAGCACTAGTACCTCTCAAAATGAGAGCTTTAGACTGCTAACCTCTGAAAGG 95

Oy 61 GGGAGCCTGTTATTTTATAG-----GGAGATG 90

Db 96 GCGAAGCTGTTATTTTATAGGGAAGATGCTTATATATATATTTTACGGAGATG 155

Oy 91 CTGTATTATTTTATCAATCCGAAATGCTGAGAAAGTATAGAAATTTGAGATG 150

Db 156 TTGTTATTTATTTATCATCTCAATATGTCAGAGAAAGTTGAAATAATTTGCAGATTG 215

Oy 151 AATCAACGTGAGAGAGAGCTTC-GAAACACTGAGACCTCGGGGCTCCT 201

Db 216 AATCAACGTGAGAGAGAGGAGCTTCAAAACACCAAGCCCTGGGGGCTCCT 267

```

1 RESULT 7
2 US-09-864-761-30590
3 Sequence 30590, Application US/09864761
4 Patent No. US20020048763A1
5 GENERAL INFORMATION:
6 APPLICANT: Penn, Sharon G.
7 APPLICANT: Rank, David R.
8 APPLICANT: Hanzel, David K.
9 APPLICANT: Chen, Wensheng
10 TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
11 FILE REFERENCE: Aecomica-X-1
12 CURRENT FILING DATE: 2001-05-23
13 PRIOR APPLICATION NUMBER: US/09/864,761
14 PRIOR FILING DATE: 2000-180,312
15 PRIOR FILING DATE: 2000-02-04
16 PRIOR APPLICATION NUMBER: US 60/207,456
17 PRIOR FILING DATE: 2000-05-26
18 PRIOR APPLICATION NUMBER: US 09/632,366
19 PRIOR FILING DATE: 2000-08-03
20 PRIOR APPLICATION NUMBER: GB 24263.6
21 PRIOR FILING DATE: 2000-10-04
22 PRIOR APPLICATION NUMBER: US 60/236,359
23 PRIOR FILING DATE: 2000-09-27
24 PRIOR APPLICATION NUMBER: PCT/US01/00666
25 PRIOR FILING DATE: 2001-01-30
26 PRIOR APPLICATION NUMBER: PCT/US01/00667
27 PRIOR FILING DATE: 2001-01-30
28 PRIOR APPLICATION NUMBER: PCT/US01/00664
29 PRIOR FILING DATE: 2001-01-30
30 PRIOR APPLICATION NUMBER: PCT/US01/00669
31 PRIOR FILING DATE: 2001-01-30
32 PRIOR APPLICATION NUMBER: PCT/US01/00665
33 PRIOR FILING DATE: 2001-01-30
34 PRIOR APPLICATION NUMBER: PCT/US01/00668
35 PRIOR FILING DATE: 2001-01-30
36 PRIOR APPLICATION NUMBER: PCT/US01/00663
37 PRIOR FILING DATE: 2001-01-30
38 PRIOR APPLICATION NUMBER: PCT/US01/00662
39 PRIOR FILING DATE: 2001-01-30
40 PRIOR APPLICATION NUMBER: PCT/US01/00661
41 PRIOR FILING DATE: 2001-01-30
42 PRIOR APPLICATION NUMBER: PCT/US01/00670
43 PRIOR FILING DATE: 2001-01-30
44 PRIOR APPLICATION NUMBER: US 60/234,687
45 PRIOR FILING DATE: 2000-09-21
46 PRIOR APPLICATION NUMBER: US 09/608,408
47 PRIOR FILING DATE: 2000-06-30
48 PRIOR APPLICATION NUMBER: US 09/774,203
49 PRIOR FILING DATE: 2001-01-29
50 NUMBER OF SEQ ID NOS: 49117
51 SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
52 SEQ ID NO 30590
53 LENGTH: 152
54 TYPE: DNA
55 ORGANISM: Homo sapiens
56 FEATURE:
57 OTHER INFORMATION: MAP TO AL022067.1

```

```
; OTHER INFORMATION: EXPRESSED IN BT474, SIGNAL = 0.95
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.1
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.3
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 7.1
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 1.2
; OTHER INFORMATION: EXPRESSED IN HBL100, SIGNAL = 1.1
; OTHER INFORMATION: NT HIT: 914759007, EVALUE 3.00e-54
; OTHER INFORMATION: EST HUMAN HIT: AUI38097.1, EVALUE 4.00e-54
; OTHER INFORMATION: SWISSPROT HIT: P10269, EVALUE 5.00e-10
US-09-864-761-30590

Query Match          48.5%; Score 97.4; DB 10; Length 152;
Best Local Similarity 90.4%; Pred. No. 1e-20;
Matches 104; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

Oy 1 CCTAGCAGCAGTAGCTCTTCAAAATCGAAGAGCTTTAGACTTGTACCGCTGAAGAAGG 60
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 29 CCTAGCAGCAGTAGCTCTTCAAAATCGAAGAGCTTTAGACTTGTACCGCTGAAGAAGG 88

Oy 61 GGAACCTGTTATTTTATAGGGAAGATGCTTTATATGTTATATCATCCGA 115
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 89 GGAAGCTGTTATTTTATAGGGAAGATGTTATATATGTTATTTATGCGGA 143

RESULT 8
US-09-864-761-7310
; Sequence 7310, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
; FILE REFERENCE: Aecm1ca-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263,6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
```

```
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Annonax Sequence Listing Engine vers. 1.1
; SEQ ID NO 7310
; LENGTH: 520
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE: MAP TO AC010135.3
; OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 1.4
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 1.5
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.7
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.6
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 1.4
US-09-864-761-7310

Query Match          44.5%; Score 89.4; DB 10; Length 520;
Best Local Similarity 70.2%; Pred. No. 5.1e-18;
Matches 134; Conservative 0; Mismatches 56; Indels 1; Gaps 1;

Oy 2 CTAGCAGCAGTAGCTCTTCAAAATCGAAGAGCTTTAGACTTGTACCGCTGAAGAAGG 61
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 41 CTAGCAGCAGTAGCTCTTCAAAATCGAAGAGCTTTAGACTTGTACCGCTGAAGAAGG 100

Oy 62 GGAACCTGTTATTTTATAGGGAAGATGCTTTATATGTTATATCATCCGAATGCTC 121
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 101 GGAACCTTCATCTCTTGGGAGAGAGAGTGTCTTTCATTAACCACTGACCAATCAT 160

Oy 122 ACTGGAAGATTAAAGAAATCGAGATGCAATACAGTACGAGAGAGCTTGGAAC 181
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 161 ACTGGAAGATTAAATTAATTAAGAGACGGA-AAAAAGTAGAAAAAGAGCTTATC 219

Oy 182 ACTGGAACCTG 192
    ||||| |||||
Db 220 TCTGAACCTG 230

RESULT 9
US-09-864-761-7501/c
; Sequence 7501, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
; TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL F
; FILE REFERENCE: Aecm1ca-X-1
; CURRENT APPLICATION NUMBER: US/09/864,761
; CURRENT FILING DATE: 2001-05-23
; PRIOR APPLICATION NUMBER: US 60/180,312
; PRIOR FILING DATE: 2000-02-04
; PRIOR APPLICATION NUMBER: US 60/207,456
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: US 09/632,366
; PRIOR FILING DATE: 2000-08-03
; PRIOR APPLICATION NUMBER: GB 24263,6
; PRIOR FILING DATE: 2000-10-04
; PRIOR APPLICATION NUMBER: US 60/236,359
; PRIOR FILING DATE: 2000-09-27
; PRIOR APPLICATION NUMBER: PCT/US01/00666
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00667
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00664
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00669
; PRIOR FILING DATE: 2001-01-30
```

```

; PRIOR APPLICATION NUMBER: PCT/US01/00665
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00668
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00663
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00662
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00661
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: PCT/US01/00670
; PRIOR FILING DATE: 2001-01-30
; PRIOR APPLICATION NUMBER: US 60/234,687
; PRIOR FILING DATE: 2000-09-21
; PRIOR APPLICATION NUMBER: US 09/608,408
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: US 09/774,203
; PRIOR FILING DATE: 2001-01-29
; NUMBER OF SEQ ID NOS: 49117
; SOFTWARE: Anomax Sequence Listing Engine vers. 1.1
; SEQ ID NO 7501
; LENGTH: 559
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: MAP TO AL139245.2
; OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 2.5
; OTHER INFORMATION: EXPRESSED IN HELA, SIGNAL = 3.9
; OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 2
; OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 3.3
; OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 3.2
; OTHER INFORMATION: EXPRESSED IN LONG, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 1.9
; OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.2
; US-09-864-761-7501

Query Match
Best Local Similarity 42.0%; Score 84.4; DB 10; Length 559;
Matches 88; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 108 AATCCGGAATCGTCAGTAAAGAAATTCGATCGAATCAACGATAGACGAG 167
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 559 AATCCAGATGTCACCGAGAAAGTTAAAGAAATTCGATCGAATCAACATAGACGAG 500
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 168 AGGAGCTTCGAACACCTGGGCGCTCCT 201
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 499 AGGAGCTTCGAACACCTGGGCGCTCCT 466
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 10
US-09-764-891-283
; Sequence 283, Application US/09764891
; Publication No. US20030077808A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PC006
; CURRENT APPLICATION NUMBER: US/09/764,891
; PRIOR APPLICATION data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 10231
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 283
; LENGTH: 960
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SITE
; LOCATION: (890)
; OTHER INFORMATION: n equals a,t,g, or c
; NAME/KEY: SITE
; LOCATION: (918)
; OTHER INFORMATION: n equals a,t,g, or c
```

```

; NAME/KEY: SITE
; LOCATION: (940)
; OTHER INFORMATION: n equals a,t,g, or c
US-09-764-891-283

Query Match
Best Local Similarity 36.5%; Score 73.4; DB 9; Length 960;
Matches 101; Conservative 1; Mismatches 47; Indels 0; Gaps 0;

QY 3 TAGCAGCAGTAGCTCTCAAAATCGAAGAGCTTAGACTTACCGCTGAAGAGGCG 62
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 383 TAGCAGCAGTAGCTCTCAAAATCGAAGAGCTTAGACTTACCGCTGAAGAGGCG 442
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 63 GAACCTGTTATTTTAAAGAGAGATGCTTTATTAATCAATCCGAATGCTCA 122
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 443 GAATTTGTTGGCTTATGATCAAAAATGTTCTTTGGTAAATCAATCAGAAAGTAC 502
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 123 CTGAGAAAGTTAAAGAAATTCGAGATCGA 151
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 503 AAGACAACATCAGACAACCTCTTAATCAA 531
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 11
US-10-001-835-58
; Sequence 58, Application US/10001835
; Patent No. US20020160387A1
; GENERAL INFORMATION:
; APPLICANT: Salceda, Susana
; APPLICANT: Macina, Roberto
; APPLICANT: Reclapon, Hervé
; APPLICANT: Caferkey, Robert
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Ovary Specific Genes and
; FILE REFERENCE: DEX-0277
; CURRENT APPLICATION NUMBER: US/10/001,835
; PRIOR FILING DATE: 2001-11-20
; PRIOR APPLICATION NUMBER: 60/249,997
; NUMBER OF SEQ ID NOS: 228
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 58
; LENGTH: 5434
; TYPE: DNA
; ORGANISM: Homo sapien
; US-10-001-835-58

Query Match
Best Local Similarity 36.5%; Score 73.4; DB 9; Length 5434;
Matches 104; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 3 TAGCAGCAGTAGCTCTCAAAATCGAAGAGCTTAGACTTACCGCTGAAGAGGCG 62
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3359 TAGCTGAGATGTCCTCCCAAAACCGCGAGCCCTGACTTACACAGCTGAAAAAGAG 3418
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 63 GAACCTGTTATTTTAAAGAGAGATGCTTTATTAATCAATCCGAATGCTCA 122
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3419 GACTCTGCATATCTTAAAGAGAGAGTGTGTTTACCTAAATCAATCTGGCTGTGT 3478
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 123 CTGAGAAAGTTAAAGAAATTCGAGATCGAATCA 157
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3479 ATGACAACATTAATAAACTCAAGATAGACCCCA 3513
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 12
US-09-864-761-24000/C
; Sequence 24000, Application US/09864761
; Patent No. US20020048763A1
; GENERAL INFORMATION:
; APPLICANT: Penn, Sharon G.
; APPLICANT: Rank, David R.
; APPLICANT: Hanzel, David K.
; APPLICANT: Chen, Wensheng
```



TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR  
FILE OF INVENTION: GENE EXPRESSION ANALYSIS BY MICROARRAY  
FILE REFERENCE: Aeomica-X-1  
CURRENT APPLICATION NUMBER: US/09/864,761  
PRIORITY FILING DATE: 2001-05-23  
PRIORITY APPLICATION NUMBER: US 60/180,312  
PRIORITY FILING DATE: 2000-02-04  
PRIORITY APPLICATION NUMBER: US 60/207,456  
PRIORITY FILING DATE: 2000-05-26  
PRIORITY APPLICATION NUMBER: US 09/632,366  
PRIORITY FILING DATE: 2000-08-03  
PRIORITY APPLICATION NUMBER: GB 24263,6  
PRIORITY FILING DATE: 2000-10-04  
PRIORITY APPLICATION NUMBER: US 60/236,359  
PRIORITY FILING DATE: 2000-09-27  
PRIORITY APPLICATION NUMBER: PCT/US01/00666  
PRIORITY FILING DATE: 2001-01-30  
PRIORITY APPLICATION NUMBER: PCT/US01/00667  
PRIORITY FILING DATE: 2001-01-30  
PRIORITY APPLICATION NUMBER: PCT/US01/00664  
PRIORITY FILING DATE: 2001-01-30  
PRIORITY APPLICATION NUMBER: PCT/US01/00669  
PRIORITY FILING DATE: 2001-01-30  
PRIORITY APPLICATION NUMBER: PCT/US01/00665  
PRIORITY FILING DATE: 2001-01-30  
PRIORITY APPLICATION NUMBER: PCT/US01/00668  
PRIORITY FILING DATE: 2001-01-30  
PRIORITY APPLICATION NUMBER: PCT/US01/00663  
PRIORITY FILING DATE: 2001-01-30  
PRIORITY APPLICATION NUMBER: PCT/US01/00662  
PRIORITY FILING DATE: 2001-01-30  
PRIORITY APPLICATION NUMBER: PCT/US01/00661  
PRIORITY FILING DATE: 2001-01-30  
PRIORITY APPLICATION NUMBER: PCT/US01/00670  
PRIORITY FILING DATE: 2001-01-30  
PRIORITY APPLICATION NUMBER: US 60/234,687  
PRIORITY FILING DATE: 2000-09-21  
PRIORITY APPLICATION NUMBER: US 09/608,408  
PRIORITY FILING DATE: 2000-06-30  
PRIORITY APPLICATION NUMBER: US 09/774,203  
PRIORITY FILING DATE: 2001-01-29  
NUMBER OF SEQ ID NOS: 49117  
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
SEQ ID NO 24000  
LENGTH: 310  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AL138934.2  
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.3  
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3  
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.5  
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 4.7  
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.9  
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.5  
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.8  
OTHER INFORMATION: SWISSPROT HIT: P04027, EVALUATE 3.00e-24  
OTHER INFORMATION: EST\_HUMAN HIT: AW847812.1, EVALUATE 1.00e-117  
US-09-864-761-24000

Query Match 35.7%; Score 71.8; DB 10; Length 310;  
Best Local Similarity 66.5%; Pred. No. 1.2e-12;  
Matches 103; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 3 TAGGAGAGTAGTCTTGAATGAAGAGCTTAGAGCTTACCGCTGAAGAGGG 62  
DB 174 TAGCTGAGTGTCTCTCAAAACGCGAGGCTTGACTTACTGCTGAAGAGAG 115  
QY 63 GAACCTGTTATTTTAGGAGAGAGAGCTTATTTATGTTATTAATCAATCCGATCGTCA 122  
DB 114 GACTCTGATATCTTAATGAAGAGAGTGTGTTTAACTTAATCAATCTGCGCTGTGT 55

QY 123 CTGAGAAAGTTAAGAAATTCGAGATCGAATACAA 157  
DB 54 ATGACAAACATATAAAAACTCAAGGATAGAGCCCAA 20

RESULT 13  
US-09-864-761-7269/c  
Sequence-7269, Application US/09864761  
Patent No. US20020048763A1  
GENERAL INFORMATION:  
APPLICANT: Penn, Sharron G.  
APPLICANT: Rank, David R.  
APPLICANT: Hanzel, David K.  
APPLICANT: Chen, Wensheng  
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL F  
FILE REFERENCE: Aeomica-X-1  
CURRENT APPLICATION NUMBER: US/09/864,761  
PRIORITY FILING DATE: 2001-05-23  
PRIORITY APPLICATION NUMBER: US 60/180,312  
PRIORITY FILING DATE: 2000-02-04  
PRIORITY APPLICATION NUMBER: US 60/207,456  
PRIORITY FILING DATE: 2000-05-26  
PRIORITY APPLICATION NUMBER: US 09/632,366  
PRIORITY FILING DATE: 2000-08-03  
PRIORITY APPLICATION NUMBER: GB 24263,6  
PRIORITY FILING DATE: 2000-10-04  
PRIORITY APPLICATION NUMBER: US 60/236,359  
PRIORITY FILING DATE: 2000-09-27  
PRIORITY APPLICATION NUMBER: PCT/US01/00666  
PRIORITY FILING DATE: 2001-01-30  
PRIORITY APPLICATION NUMBER: PCT/US01/00667  
PRIORITY FILING DATE: 2001-01-30  
PRIORITY APPLICATION NUMBER: PCT/US01/00664  
PRIORITY FILING DATE: 2001-01-30  
PRIORITY APPLICATION NUMBER: PCT/US01/00669  
PRIORITY FILING DATE: 2001-01-30  
PRIORITY APPLICATION NUMBER: PCT/US01/00665  
PRIORITY FILING DATE: 2001-01-30  
PRIORITY APPLICATION NUMBER: PCT/US01/00668  
PRIORITY FILING DATE: 2001-01-30  
PRIORITY APPLICATION NUMBER: PCT/US01/00663  
PRIORITY FILING DATE: 2001-01-30  
PRIORITY APPLICATION NUMBER: PCT/US01/00662  
PRIORITY FILING DATE: 2001-01-30  
PRIORITY APPLICATION NUMBER: PCT/US01/00661  
PRIORITY FILING DATE: 2001-01-30  
PRIORITY APPLICATION NUMBER: PCT/US01/00670  
PRIORITY FILING DATE: 2001-01-30  
PRIORITY APPLICATION NUMBER: US 60/234,687  
PRIORITY FILING DATE: 2000-09-21  
PRIORITY APPLICATION NUMBER: US 09/608,408  
PRIORITY FILING DATE: 2000-06-30  
PRIORITY APPLICATION NUMBER: US 09/774,203  
PRIORITY FILING DATE: 2001-01-29  
NUMBER OF SEQ ID NOS: 49117  
SOFTWARE: Anomax Sequence Listing Engine vers. 1.1  
SEQ ID NO 7269  
LENGTH: 556  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
OTHER INFORMATION: MAP TO AL138934.2  
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.3  
OTHER INFORMATION: EXPRESSED IN ADULT LIVER, SIGNAL = 3  
OTHER INFORMATION: EXPRESSED IN BRAIN, SIGNAL = 2.5  
OTHER INFORMATION: EXPRESSED IN BONE MARROW, SIGNAL = 4.7  
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 2.9  
OTHER INFORMATION: EXPRESSED IN HEART, SIGNAL = 1.5  
OTHER INFORMATION: EXPRESSED IN FETAL LIVER, SIGNAL = 2.8  
US-09-864-761-7269

Query Match 35.7%; Score 71.8; DB 10; Length 556;

Best Local Similarity 66.5%; Pred. No. 1.5e-12;  
Matches 103; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 3 TGGACGACATAGTCCTTCAAAATCGAAGAGCTTTAGCTTGTACACCGCTGAAGAGGG 62  
DB 209 TAGCTGCACTTGTCTCCCAAAACCGCCGAGCCTTACTTACTCAATGCTGAAGAGAG 150  
QY 63 GAACCTGTTATTTTATGAGGAGAAATGCTTATTTATGTTAATCAATCCGAAATCGTCA 122  
DB 149 GACTCTGCATATTTCTTAATAGAGAGAGTGTGTTTACTTAATCAATCTGGCCTGTGT 90  
QY 123 CTGAGAAAGTTAAAGAAATTCGAGATCGAATACAA 157  
DB 89 ATGACAACTATAAAAACTCAAGATAGAGCCCAA 55

RESULT 14  
US-09-729-674-125  
; Sequence 125, Application US/09729674  
; Patent No. US20010039335A1  
; GENERAL INFORMATION:  
; APPLICANT: Jacobs, Kenneth  
; APPLICANT: McCoy, John M.  
; APPLICANT: Lavallee, Edward R.  
; APPLICANT: Collins-Racie, Lisa A.  
; APPLICANT: Evans, Cheryl  
; APPLICANT: Merberg, David  
; APPLICANT: Treacy, Maurice  
; APPLICANT: Agostino, Michael J.  
; APPLICANT: Steindinger II, Robert J.  
; APPLICANT: Spaulding, Vikki  
; APPLICANT: Wong, Gordon G.  
; APPLICANT: Clark, Hilary  
; APPLICANT: Fechtel, Kim  
; APPLICANT: Genetics Institute, Inc.  
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM  
; FILE REFERENCE: 6055-64X  
; CURRENT APPLICATION NUMBER: US/09/729,674  
; CURRENT FILING DATE: 2000-12-04  
; PRIOR APPLICATION NUMBER: 09/539,330  
; PRIOR FILING DATE: 2000-03-30  
; NUMBER OF SEQ ID NOS: 283  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 125  
; LENGTH: 1783  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-729-674-125

Query Match 32.7%; Score 65.8; DB 10; Length 1783;  
Best Local Similarity 68.4%; Pred. No. 1.7e-10;  
Matches 91; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

QY 1 CCTAGCAGCAGTAGTCCTTCAAAATCGAAGAGCTTTAGCTTGTACACCGCTGAAGAGG 60  
DB 911 CCTCGTGGAGTCATCTTCAAAATCGAGAGTCCCTTAACCACTGAGAAAGG 970  
QY 61 GGAACCTGTTATTTTATGAGGAGAAATGCTTATTTATGTTAATCAATCCGAAATCGT 120  
DB 971 GGGTACCTCATATTAATCTCCAGAGAAATGCTGTTTGTGTTAATGAATCTGGCAATGT 1030  
QY 121 CACTGAGAAAGTT 133  
DB 1031 TCATATCGCAGTT 1043

RESULT 15  
US-09-739-254-12  
; Sequence 12, Application US/09739254  
; Patent No. US20010021700A1  
; GENERAL INFORMATION:  
; APPLICANT: Rosen et al.  
; TITLE OF INVENTION: 49 Human Secreted Proteins

; FILE REFERENCE: P2032P1  
; CURRENT APPLICATION NUMBER: US/09/739,254  
; CURRENT FILING DATE: 2000-12-19  
; EARLIER APPLICATION NUMBER: 09/511,554  
; EARLIER FILING DATE: 2000-02-23  
; EARLIER APPLICATION NUMBER: PCT/US99/19330  
; EARLIER FILING DATE: 1999-08-24  
; EARLIER APPLICATION NUMBER: 60/097,917  
; EARLIER FILING DATE: 1998-08-25  
; EARLIER APPLICATION NUMBER: 60/098,634  
; EARLIER FILING DATE: 1998-08-31  
; NUMBER OF SEQ ID NOS: 170  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 12  
; LENGTH: 2342  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-739-254-12

Query Match 32.7%; Score 65.8; DB 10; Length 2342;  
Best Local Similarity 68.4%; Pred. No. 1.9e-10;  
Matches 91; Conservative 0; Mismatches 42; Indels 0; Gaps 0;

QY 1 CCTAGCAGCAGTAGTCCTTCAAAATCGAAGAGCTTTAGCTTGTACACCGCTGAAGAGG 60  
DB 1483 CCTCGTGGAGTCATCTTCAAAATCGAGAGTCCCTTAACCACTGAGAAAGG 1542  
QY 61 GGAACCTGTTATTTTATGAGGAGAAATGCTTATTTATGTTAATCAATCCGAAATCGT 120  
DB 1543 GGGTACCTCATATTAATCTCCAGAGAAATGCTGTTTGTGTTAATGAATCTGGCAATGT 1602  
QY 121 CACTGAGAAAGTT 133  
DB 1603 TCATATCGCAGTT 1615

Search completed: May 2, 2003, 18:41:58  
Job time : 69.3333 secs